Genealogia sztucznej inteligencji w ujęciu Dennisa Yi Tenena: od struktur narracyjnych i probabilistyki do politycznej ekonomii wiedzy i etyki odpowiedzialności.

AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Tekst analizuje ewolucję sztucznej inteligencji poprzez pryzmat struktur narracyjnych i lingwistyki formalnej. Autor śledzi przejście od rosyjskiego formalizmu Władimira Proppa i jego morfologii bajki do współczesnych systemów obliczeniowych, takich jak TALE-SPIN. Kluczowym punktem jest zmiana epistemologiczna: traktowanie narracji nie jako utworu literackiego, lecz jako systemu reguł, funkcji i relacji, na których można wykonywać operacje matematyczne. Wpływ strukturalizmu Ferdinanda de Saussure'a oraz Claude'a Lévi-Straussa ukazuje dążenie do odkrycia uniwersalnej „gramatyki kultury". To systemowe ujęcie języka i mitu stało się fundamentem dla późniejszego modelowania zdarzeń w AI, umożliwiając oddzielenie zmiennej treści od stabilnej struktury funkcjonalnej.

The text analyzes the evolution of artificial intelligence through the prism of narrative structures and formal linguistics. The author traces the transition from Vladimir Propp's Russian formalism and his morphology of the folk tale to modern computational systems such as TALE-SPIN. A key point is the epistemological shift: treating narration not as a literary work, but as a system of rules, functions, and relations upon which mathematical operations can be performed. The influence of Ferdinand de Saussure's structuralism and Claude Lévi-Strauss reveals the drive to discover a universal 'grammar of culture.' This systemic approach to language and myth became

the foundation for later event modeling in AI, enabling the separation of variable content from stable functional structure.

Artykuł stanowi krytyczną analizę natury współczesnej sztucznej inteligencji, przeciwstawiając powszechną tendencję do personifikowania maszyn rygorystycznemu ujęciu genealogicznemu i filozoficznemu. Autor dowodzi, że AI nie jest autonomicznym podmiotem obdarzonym rozumieniem czy wolą, lecz technologiczną kondensacją ludzkiej pracy, statystycznych regularności języka oraz sformalizowanych struktur narracyjnych, których korzenie sięgają strukturalizmu i teorii informacji. Główna teza zakłada, że przypisywanie systemom AI cech sprawstwa moralnego lub intelektualnego jest nie tylko błędem kategorialnym, ale przede wszystkim niebezpiecznym zabiegiem retorycznym. Taka antropomorfizacja przesłania rzeczywiste relacje władzy i ekonomiczne uwarunkowania produkcji wiedzy, pozwalając instytucjom oraz korporacjom na rozmycie odpowiedzialności prawnej i etycznej. Tekst przekształca debatę o „świadomości maszyn" w polityczną ekonomię poznania, w której kluczowym pytaniem nie jest to, czy AI myśli, lecz kto kontroluje infrastrukturę tego procesu i kto ponosi konsekwencje decyzji podejmowanych przez algorytmy.

SŁOWA KLUCZOWE

genealogia sztucznej inteligencji | struktury narracyjne | probabilistyka języka | polityczna ekonomia wiedzy | etyka odpowiedzialności | morfologia bajki | gramatyka generatywna | systemy probabilistyczne | TALE-SPIN | reprezentacja stanu świata | sprawstwo funkcjonalne | moral agency | moral patiency | luka odpowiedzialności | modelowanie narracji

Narracja jako system reguł i funkcji

Airplane Stories – od morfologii bajki do obliczeniowej reprezentacji zdarzeń. Rozdział ten zajmuje w genealogii Dennisa Yi Tenena miejsce szczególne; to tutaj wcześniejsze dzieje kombinatoryki, szablonu i mechanicznej reprezentacji splatają się z dwudziestowieczną lingwistyką formalną, narratologią oraz pierwszymi programami komputerowymi. Te ostatnie nie dążyły jedynie do

składania poprawnych zdań, lecz próbowały modelować zdarzenia jako uporządkowane ciągi celów, przeszkód, działań i skutków. Materiał źródłowy prowadzi tę linię od Władimira Proppa, przez Noama Chomsky'ego i Victora Yngvego, aż po program TALE-SPIN Jamesa Meehana oraz systemy przetwarzania narracji technicznych i lotniczych. Istotą tej sekwencji nie jest jednak prosta opowieść o „uczeniu komputera opowiadania bajek". Znacznie ważniejsza jest przemiana epistemologiczna: narracja przestaje być traktowana wyłącznie jako niepowtarzalny utwór literacki, a zaczyna być badana jako struktura posiadająca składniki, relacje, ograniczenia i reguły transformacji. Dopiero taka zmiana umożliwiła późniejsze przedstawienie opowieści w formie, na której można wykonywać operacje obliczeniowe.

Punktem wyjścia jest rosyjski formalizm i „Morfologia bajki" Władimira Proppa z 1928 roku. Propp nie skupiał się na psychologicznej głębi bohaterów ani indywidualnym stylu bajarza. Interesowało go, czy pod ogromną różnorodnością rosyjskich bajek magicznych kryje się powtarzalna organizacja działań. Kluczową kategorią stała się „funkcja" – działanie postaci zdefiniowane przez jego znaczenie dla przebiegu fabuły. Badania doprowadziły go do wyróżnienia trzydziestu jeden podstawowych funkcji oraz ograniczonego repertuaru ról. Choć nie każda bajka zawiera wszystkie te elementy, ich kolejność wykazuje znacznie większą regularność, niż sugerowałaby powierzchniowa różnorodność imion, rekwizytów i dekoracji. Współczesna teoria narracji uznaje Proppa za jednego z głównych prekursorów strukturalistycznej narratologii, gdyż przesunął on punkt ciężkości z jednostkowego tekstu na formalny opis systemu możliwości narracyjnych. Ruch ten przyniósł daleko idące konsekwencje metodologiczne.

Gdy bohater jednej bajki wyrusza po zaczarowany pierścień, a inny po wodę życia, analiza powierzchniowa dostrzega dwa różne przedmioty i dwie odmienne historie. Morfologia Proppa pozwala natomiast potraktować oba epizody jako realizacje tej samej funkcji fabularnej. Zmienna treść zostaje tu oddzielona od relacyjnej pozycji elementu w strukturze. Nie jest to jeszcze algorytm generowania opowieści, lecz warunek umożliwiający późniejszą algorytmizację: aby system komputerowy mógł operować na narracji, należy najpierw określić, które właściwości zdarzenia są przypadkowe, a które pełnią stabilną funkcję w strukturze. Propp wykonuje dla bajki gest analogiczny do tego, który gramatyka stosuje wobec zdania: przechodzi od konkretnych realizacji do abstrakcyjnego systemu ograniczeń. Nieprzypadkowo jego praca została później włączona do szerokiej konstelacji strukturalizmu.

Strukturalizm jako fundament systemowego ujęcia języka i narracji

Ferdinand de Saussure dokonał jednego z zasadniczych przesunięć w nowoczesnym językoznawstwie, odróżniając langue – społeczny system języka – od parole, czyli jego konkretnych użyć, oraz przeciwstawiając analizę synchroniczną badaniu przemian diachronicznych. Kluczowa okazała się koncepcja języka jako systemu różnic, w którym wartość jednostki nie wynika z jej izolowanej substancji, lecz z miejsca zajmowanego względem innych elementów. W późniejszej recepcji to właśnie relacyjne ujęcie znaczenia stało się bazą strukturalizmu w antropologii, semiotyce i teorii literatury. Cambridge Companion to Saussure podkreśla, że projekt ten zmierzał do nadania językoznawstwu rygoru naukowego poprzez systematyczne rozróżnienie między langue a parole, synchronią a diachronią oraz różnymi porządkami relacji znakowych. Podobną logikę Claude Lévi-Strauss przeniósł do antropologii.

Mit w jego ujęciu nie był jedynie opowieścią przekazującą treść kulturową, lecz strukturą relacji, w której znaczenie ujawnia się poprzez transformacje i opozycje. Natura i kultura, surowe i gotowane, życie i śmierć czy tożsamość i różnica tworzyły systemy przeciwstawień organizujące mityczne warianty. Ambicją strukturalistów było wykazanie, że powierzchniowo odmienne opowieści mogą być transformacjami wspólnej struktury relacyjnej. Nie oznaczało to, że całą kulturę da się sprowadzić do kilku binarnych formuł; późniejsza krytyka, od Derridy po antropologię interpretatywną, wskazywała na ograniczenia takiego formalizmu. Sam fakt, że Lévi-Strauss próbował wyrażać relacje mityczne za pomocą formuł, świadczy jednak o silnej dwudziestowiecznej nadziei na odkrycie "gramatyki kultury" – niewidocznego systemu umożliwiającego tworzenie wielości powierzchniowych realizacji.

Narratologia strukturalistyczna może być traktowana jako jeden z intelektualnych przodków komputerowej generacji narracji, choć nie należy widzieć w tej relacji prostego łańcucha bezpośrednich wpływów. Tzvetan Todorov, Roland Barthes i Gérard Genette rozwinęli aparat pozwalający odróżnić zdarzenia tworzące historię od sposobu ich dyskursywnego przedstawienia oraz analizować porządek, czas trwania, częstotliwość, perspektywę i głos narracyjny. Narratologia dążyła do opisu systemu reguł odpowiedzialnych za produkcję i rozpoznawanie struktur narracyjnych, podobnie jak językoznawstwo opisuje system warunkujący tworzenie zdań. Cambridge History of Literary Criticism ujmuje klasyczną narratologię właśnie jako strukturalistyczne poszukiwanie "narrative langue" – systemu umożliwiającego wielość konkretnych narracyjnych paroles.

Relacja między strukturalizmem a Chomskym wymaga jednak szczególnej precyzji, ponieważ materiał źródłowy upraszcza historię, przedstawiając gramatykę generatywną niemal jako dalszy etap strukturalizmu i jako "test", w którym regułę sprawdza się poprzez zdolność maszyny do odtworzenia języka. Historycznie sprawa jest bardziej złożona.

Rozdźwięk między regułami gramatycznymi a statystyką języka

Chomsky sprzeciwiał się założeniom amerykańskiego strukturalizmu dystrybucyjnego i behawioryzmu, postulując teorię zdolną wyjaśnić generatywny charakter kompetencji językowej — czyli zdolność do tworzenia i rozumienia potencjalnie nieograniczonej liczby nowych zdań na podstawie skończonego zbioru zasad. Jego słynne zdanie „Colorless green ideas sleep furiously" nie miało przede wszystkim dowodzić, że maszyna nie zna świata, jak sugerują niektóre współczesne analogie do AI. W „Syntactic Structures" pełniło ono precyzyjniejszą funkcję argumentacyjną: miało wykazać, że gramatyczności nie można utożsamić ani z sensownością, ani z wysokim prawdopodobieństwem wystąpienia danego ciągu w korpusie językowym. Chomsky wskazywał, że zdanie może być syntaktycznie poprawne mimo semantycznej dziwności, podczas gdy jego prosty odpowiednik z odwróconym szykiem nie jest już akceptowalnym zdaniem w języku angielskim.

To rozróżnienie ma fundamentalne znaczenie dla historii AI, gdyż ujawnia dwie konkurencyjne tradycje modelowania języka. Pierwsza poszukuje jawnych reguł generatywnych określających przestrzeń struktur dopuszczalnych; druga rozwija podejście probabilistyczne, w którym kluczowe staje się przewidywanie elementów na podstawie ich statystycznych zależności.

Współczesnych LLM nie należy zatem projektować „wstecz" do Chomsky'ego. Wręcz przeciwnie — jego klasyczny przykład był częściowo wymierzony w przekonanie, że samo statystyczne przybliżenie korpusu wystarczy do wyjaśnienia gramatyczności. Dzisiejsze modele językowe stanowią niezwykle interesujący kontrapunkt dla tego argumentu; ogromne systemy probabilistyczne nauczyły się reprezentować liczne właściwości składni i semantyki bez potrzeby ręcznego zapisywania gramatyki transformacyjnej. Nie rozstrzyga to filozoficznego sporu o naturę języka, ale pokazuje, jak rozwój techniki zatarł podział między formalną regułą a statystyczną regularnością.

Wczesne doświadczenia Victora Yngvego dobrze ilustrują przejście od lingwistyki formalnej do realnej generacji komputerowej. W pracy „Random Generation of English Sentences" z 1961 roku Yngve konstruował system reguł umożliwiających automatyczne wytwarzanie angielskich zdań.

Badania te wyrastały z prac nad tłumaczeniem maszynowym, gdzie niezbędny był nie tylko mechanizm analizy tekstu wejściowego, ale i zdolność wygenerowania poprawnej wypowiedzi w języku docelowym. Zachowana praca konferencyjna potwierdza, że eksperyment Yngvego należał do pierwszych systematycznych prób komputerowej generacji zdań. Współczesne rekonstrukcje wskazują, że do modelowania początkowych fragmentów z dziecięcej książki „The Little Train" wykorzystał on zestaw reguł rekursywnych, pozwalających na tworzenie pokrewnych wariantów.

Znaczenie tych eksperymentów nie tkwiło w ich literackiej jakości, lecz w samej możliwości przejścia od opisu języka do procedury generującej jego egzemplarze. Teoria stawała się wykonawcza.

Jeśli gramatyka rzeczywiście opisuje zasady systemu, można próbować zakodować ją tak, aby komputer wytwarzał struktury zgodne z tym opisem. W filozofii nauki jest to interesujący przykład modelu generatywnego w sensie metodologicznym: wyjaśnienie zyskuje na sile, gdy na podstawie postulowanego mechanizmu można odtworzyć klasę obserwowanych zjawisk. Nie jest to oczywiście dowód na to, że komputer generuje zdania w sposób identyczny z ludzkim mózgiem, lecz test adekwatności formalnego modelu wobec wybranego poziomu struktury językowej.

Narracja jako symulacja stanu świata a nie generowanie tekstu

Przejście od pojedynczego zdania do opowieści wymagało jednak zasadniczego wzbogacenia reprezentacji. O ile zdanie można wygenerować w oparciu o reguły składniowe, nie posiadając modelu świata, o tyle narracja wymusza utrzymanie spójnych relacji między zdarzeniami, postaciami, celami a konsekwencjami. Jeśli bohater dąży do zdobycia wody, kolejne działania muszą być z tym celem powiązane; jeśli przedmiot został przeniesiony, następna scena nie może bez wyjaśnienia zakładać jego poprzedniego położenia; jeśli jedna postać dysponuje wiedzą, której druga nie posiada, ich późniejsze zachowania powinny respektować tę asymetrię. Generator narracji potrzebuje zatem nie tylko gramatyki powierzchniowej, lecz reprezentacji stanu świata oraz teorii działania.

Na tym tle TALE-SPIN Jamesa Meehana stanowi istotny etap rozwoju sztucznej inteligencji narracyjnej. W pracy „Using Planning Structures to Generate Stories" Meehan opisał program tworzący historie za pomocą struktur planowania, będących częścią wiedzy systemu o świecie. Reprezentowały one cele oraz metody ich osiągania; działanie programu zmieniało stan modelowanego świata, wprowadzało postaci i przeszkody, co prowadziło do dalszych konsekwencji.

Narracja była następnie generowana z reprezentacji Conceptual Dependency i przekładana na język angielski przez osobny komponent językowy. Program nie „wymyślał" więc opowieści poprzez losowe dobieranie zdań. Symulował on działania postaci, których zachowanie wynikało z celów, stanów i procedur rozwiązywania problemów – dopiero wybrane zdarzenia tej symulacji stawały się materiałem literackim. To rozróżnienie między światem opowieści a tekstem opowieści jest naukowo kluczowe. Meehan wyraźnie zaznaczał, że w procesie rozwiązywania problemu zachodzi więcej zdarzeń, niż narrator musi wypowiedzieć. Opowiadanie jest zatem selekcją elementów modelu świata, a nie jego kompletnym przepisaniem. TALE-SPIN pozwala połączyć informatykę z klasycznym rozróżnieniem narratologicznym pomiędzy story i discourse: sekwencja zdarzeń modelowanych przez system to co innego niż ich językowa prezentacja. Generator opowieści potrzebuje zatem dwóch teorii: tej, co mogłoby się wydarzyć, oraz tej, co z tych zdarzeń warto przekazać i w jakiej kolejności. Sam Meehan przyznawał, że jego system znacznie lepiej radził sobie z pierwszym problemem niż z subtelnymi kwestiami stylu.

TALE-SPIN należy osadzić w szerszej tradycji kognitywnej Rogera Schanka i Roberta Abelsona. Ich teoria skryptów, planów, celów i struktur wiedzy wynikała z przekonania, że rozumienie opowieści wymaga czegoś więcej niż syntaktycznego przetworzenia zdań. Człowiek rekonstruuje niedopowiedziane związki przyczynowe dzięki wiedzy o typowych sytuacjach społecznych. Skrypt restauracji pozwala zrozumieć, dlaczego po zamówieniu posiłku pojawia się rachunek, nawet jeśli tekst nie wyjaśnia norm instytucjonalnych stojących za tym ciągiem zdarzeń. W sztucznej inteligencji skrypt stał się próbą reprezentacji wiedzy zdroworozsądkowej niezbędnej do interpretacji tekstu. Książka Schanka i Abelsona z 1977 roku jest klasyką tego nurtu, a struktury planowania Meehana wyrastały właśnie z tego środowiska intelektualnego.

Ten punkt pozwala lepiej zrozumieć Tenenowskie „Airplane Stories". Materiał źródłowy opisuje techniczne raporty lotnicze jako sformalizowane narracje, w których awaria, reakcja załogi, zmiana planu i zakończenie zdarzenia układają się w strukturę przypominającą opowieść o celu, zakłóceniu i rozwiązaniu. Sens analogii nie polega na twierdzeniu, że raport o awarii silnika jest literaturą piękną, lecz na tym, że zarówno bajka, jak i raport techniczny mogą posiadać temporalnie i przyczynowo zorganizowaną strukturę zdarzeń. Samolot ma dotrzeć do celu, pojawia się zakłócenie, aktorzy podejmują działania naprawcze, plan ulega modyfikacji, a stan końcowy jest oceniany względem pierwotnego celu. Dla systemu reprezentującego wiedzę taki ciąg można opisać za pomocą tych samych kategorii: agent, cel, stan, zdarzenie, przeszkoda, działanie i skutek.

Warto jednak rozdzielić dwie kwestie, które materiał źródłowy zbliża do siebie bardziej, niż pozwala na to dokumentacja. Tenenowska narracja wiąże historię „opowieści samolotowych" z pracami Petera Clarka i środowiskiem Boeinga, przywołując bazę FIDES jako przykład automatycznych

narracji o incydentach. Tymczasem niezależnie zweryfikowana publikacja Clarka i Phila Harrisona z 2008 roku dotyczy systemu NLP Boeinga o nazwie BLUE, lecz opisuje przede wszystkim pipeline przekształcający tekst w reprezentację logiczną: parser, generator formy logicznej, rozpoznawanie ról semantycznych, koreferencje i mapowanie na ontologię.

Niewystarczalność formy wobec głębi znaczenia

Autorzy podkreślają, że przejście od powierzchni tekstu do reprezentacji umożliwiającej poprawne wnioskowanie jest trudne nawet w przypadku pozornie prostych zdań. Dlatego szerszą Tenenowską figurę "bajek pisanych przez samoloty" należy traktować jako interpretacyjną syntezę historii generowania i analizy narracji technicznych, a nie jako prosty opis działania systemu BLUE. Co znamienne, prace tego nurtu powstawały w instytucjonalnym otoczeniu finansowania wojskowego i badań nad systemami eksperckimi. Artykuł Meehana o TALE-SPIN wprost wskazuje na finansowanie przez Advanced Research Projects Agency Departamentu Obrony oraz nadzór Office of Naval Research. Fakt ten nie dowodzi oczywiście, że generatory bajek były projektami militarnymi w wąskim sensie, ani że całą sztuczną inteligencję można redukować do potrzeb wojskowych. Pokazuje natomiast coś istotniejszego dla socjologii nauki: rozwój metod reprezentowania wiedzy, planowania, języka naturalnego i automatycznego wnioskowania był częścią większego ekosystemu instytucjonalnego, w którym badania podstawowe, uniwersytety, przemysł i państwowe finansowanie obronne pozostawały silnie splecione. Historia idei technicznych jest więc również historią infrastruktury finansującej pewne pytania i pomijającej inne.

Z perspektywy filozofii nauki najważniejszym osiągnięciem tej epoki było przesunięcie od tekstu jako ciągu znaków ku tekstowi rozumianemu jako powierzchniowa manifestacja bardziej złożonego modelu. Propp poszukiwał morfologii działań, Chomsky formalizował generatywną strukturę składni, a Yngve dowodził, że reguły te mogą zostać zaimplementowane komputerowo. Schank i Abelson próbowali modelować wiedzę o typowych sekwencjach działań, Meehan konstruował symulowany świat postaci posiadających cele, zaś Clark i współpracownicy usiłowali przekształcać językową powierzchnię w reprezentacje umożliwiające wnioskowanie. Wszystkie te przedsięwzięcia różnią się metodologicznie, lecz łączy je przekonanie, że sama powierzchnia wypowiedzi nie wystarcza: konieczna jest rekonstrukcja struktury generującej lub organizującej jej sens.

W tym miejscu ujawnia się granica strukturalistycznego marzenia. Opowieści nie są wyłącznie układami formalnymi, ponieważ ich rozumienie zależy od ogromnej ilości wiedzy niewypowiedzianej. Zdanie "pilot zawrócił do bramki" jest zrozumiałe tylko dla podmiotu

dysponującego wiedzą o lotnisku, kołowaniu, obowiązkach kapitana, bezpieczeństwie pasażerów i celowości transportu lotniczego. Wiedza ta nie zawiera się w samej składni zdania. Współczesna semantyka komputerowa nadal zmaga się z tym problemem: reprezentacja może zachowywać coraz więcej relacji, lecz zawsze pozostaje pytanie, ile świata trzeba odtworzyć, aby poprawnie zinterpretować nawet krótką wypowiedź. Clark i Harrison piszą o spektrum możliwych znaczeń "reprezentacji semantycznej" — od struktur zachowujących wiele cech składniowych po znacznie głębsze reprezentacje przeznaczone bezpośrednio do wnioskowania. Im bardziej chcemy ujawnić inferencyjne konsekwencje tekstu, tym trudniejsze staje się samo skonstruowanie reprezentacji.

Z tego powodu historia Airplane Stories powinna być odczytywana nie jako triumf formy nad znaczeniem, lecz jako odkrywanie kolejnych warstw potrzebnych do przejścia od formy do znaczenia. Propp ujawnia regularność narracyjną, lecz jego trzydzieści jeden funkcji nie wyjaśnia całej semantyki każdej bajki. Gramatyka generatywna opisuje warunki strukturalne, lecz nie rozwiązuje automatycznie problemu odniesienia. Skrypt reprezentuje typową sytuację, ale może zawodzić w przypadku nietypowym. Planowanie pozwala symulować racjonalne dążenie do celu, lecz ludzka motywacja nie zawsze jest racjonalna, przejrzysta ani stabilna.

Przejście od reguł do statystycznych regularności języka

Ontologia porządkuje relacje, lecz sama pozostaje konstrukcją teoretyczną, wymagającą decyzji co do tego, jakie byty i zależności uznamy za istotne. Każdy sukces formalizacji odsłania zatem kolejną warstwę wiedzy, która wciąż oczekuje na sformalizowanie.

Rozdział Tenena przygotowuje grunt pod następny zasadniczy zwrot. W Airplane Stories język jest wciąż ujmowany głównie przez reguły, funkcje, skrypty i reprezentacje wiedzy. W kolejnej fazie kluczową rolę zyska inna odpowiedź na problem języka: zamiast ręcznie opisywać jego strukturę i wiedzę o świecie, można badać regularności statystyczne obecne w samych tekstach. Andriej Markow dokonał tego jeszcze przed erą komputerów elektronicznych, traktując „Eugeniusza Oniegina” nie jako dzieło wymagające hermeneutycznego odczytania, lecz jako sekwencję elementów pozostających w probabilistycznych zależnościach. Claude Shannon wykorzysta później podobną logikę w matematycznej teorii komunikacji, a językoznawstwo dystrybucyjne i wyszukiwanie informacji zaczną nadawać znaczenie statystycznemu sąsiedztwu słów.

Przejście to będzie metodologicznie radykalne. O ile w Airplane Stories komputer otrzymuje rozbudowaną teorię zdarzenia, celu, agenta czy planu, o tyle w świecie otwartym przez Markowa i

Shannona system może osiągać użyteczne rezultaty bez explicite zapisanej teorii tych pojęć, wykorzystując jedynie regularności wynikające z rozkładu znaków i słów. Historia sztucznej inteligencji przesuwa się zatem od pytania „jakie reguły generują poprawną strukturę?" ku pytaniu „jakie prawdopodobieństwo ma następny element, jeśli znamy to, co pojawiło się wcześniej?". To przesunięcie nie eliminuje problemów semantycznych odkrytych przez strukturalizm i klasyczną AI, lecz przenosi je na zupełnie inny poziom reprezentacji.

Właśnie tam rozpocznie się Markov's Pushkin – od probabilistyki zależnych zdarzeń do statystycznej geometrii języka. Przejście to oznacza zmianę nie tylko techniczną, lecz przede wszystkim epistemologiczną. W strukturalizmie, gramatyce generatywnej i klasycznej sztucznej inteligencji badacz usiłował odkryć lub jawnie zaprojektować reguły organizujące język, narrację i wiedzę o świecie. W podejściu probabilistycznym punkt ciężkości zostaje przesunięty: system nie musi posiadać kompletnej teorii zdania, bohatera czy intencji, jeżeli potrafi wystarczająco dobrze modelować regularności pojawiające się w obserwowanych sekwencjach. Analiza „Eugeniusza Oniegina" przeprowadzona przez Andrieja Markowa stanowi symboliczny moment tej transformacji, prowadząc następnie ku Shannonowskiej teorii informacji, językoznawstwu dystrybucyjnemu Firtha, przestrzeniom wektorowym Saltona i statystycznemu ważeniu terminów Karen Spärck Jones. Sens tej genealogii nie polega na twierdzeniu, że współczesny duży model językowy jest jedynie rozbudowanym łańcuchem Markowa.

Łańcuchy Markowa jako model statystycznej zależności języka

Takie utożsamienie byłoby matematycznie błędne. Chodzi o znacznie głębszą zmianę paradygmatu: język może stać się przedmiotem naukowego modelowania nawet wtedy, gdy badacz czasowo zawiesza pytanie o jego znaczenie i koncentruje się na statystycznych zależnościach między obserwowalnymi elementami. Aby w pełni uchwycić sens eksperymentu Markowa, należy osadzić go w historii rachunku prawdopodobieństwa. Klasyczne zastosowania probabilistyki — od Pascala i Fermata, przez Bernoulliego, aż po Laplace'a — skupiały się głównie na sytuacjach, w których zdarzenia można było analizować jako niezależne lub przynajmniej tak je idealizować. Losowanie kuli z urny, rzut monetą czy wynik gry hazardowej sprzyjały budowaniu modeli, w których poprzednie zdarzenie nie wpływało na prawdopodobieństwo następnego.

Tymczasem ogromna część rzeczywistych procesów działa inaczej. Stan gospodarki zależy od stanu poprzedniego, choroba rozwija się w czasie, zachowanie populacji wynika z wcześniejszej struktury, pogoda wykazuje ciągłość, a język jest wręcz podręcznikowym przykładem zależności sekwencyjnej:

po określonych literach, morfemach i słowach pewne kontynuacje są bardziej prawdopodobne niż inne. Markow, związany z petersburską szkołą probabilistyki Pafnutija Czebyszewa, podjął problem granicznych praw rachunku prawdopodobieństwa w warunkach zależności. Współczesne rekonstrukcje historyczne podkreślają, że jego łańcuchy wyrosły z wewnętrznych potrzeb samej teorii prawdopodobieństwa — zwłaszcza z pytania o to, w jakim zakresie prawo wielkich liczb można rozszerzyć na sekwencje zdarzeń, które nie są niezależne. Wykorzystanie Puszkina nie było ekscentrycznym kaprysem matematyka, który postanowił sprofanować poezję liczeniem liter. Tekst literacki oferował Markowowi długą, realną sekwencję zdarzeń, w której zależność była intuicyjnie oczywista, a jednocześnie możliwa do ilościowego pomiaru. W odczycie z 23 stycznia 1913 roku w Petersburskiej Akademii Nauk Markow analizował fragment Eugeniusza Oniegina obejmujący 20 000 liter (z pierwszego rozdziału i szesnastu strof drugiego), klasyfikując każdą z nich jako samogłoskę lub spółgłoskę, co pozwoliło uzyskać dwustanowy opis sekwencji.

Oryginalna praca została wiele dekad później przełożona i opublikowana w Science in Context, dzięki czemu możemy precyzyjnie odróżnić rzeczywisty eksperyment od narosłej wokół niego legendy. Materiał źródłowy poprawnie zachowuje zasadnicze elementy tego opisu: 20 000 znaków, usunięcie interpunkcji, redukcję liter do dwóch klas i badanie "powiązanych prób". Z perspektywy teorii procesów stochastycznych kluczowe było pytanie, czy prawdopodobieństwo następnego stanu zależy od poprzedniego. Markow wykazał, że w analizowanym tekście samogłoski i spółgłoski nie występują jak niezależne rzuty monetą. Po samogłosce następowała spółgłoska z wysokim prawdopodobieństwem (ok. 0,872), podczas gdy po spółgłosce samogłoska pojawiała się z prawdopodobieństwem ok. 0,663. Odpowiednio przejście samogłoska-samogłoska miało prawdopodobieństwo ok. 0,128, a spółgłoska-spółgłoska ok. 0,337.

Całą dynamikę można opisać za pomocą macierzy przejścia, w której każdy stan posiada rozkład prawdopodobieństwa kolejnego stanu. Współczesna rekonstrukcja matematyczna wskazuje także rozkład stacjonarny (ok. 43,2% samogłosek i 56,8% spółgłosek), zgodny z empirycznymi proporcjami w materiale. Nie jest to jednak jedynie historyczna ciekawostka.

Markow zademonstrował ogólną ideę: proces może być jednocześnie losowy i strukturalnie zależny. "Losowość" w sensie probabilistycznym nie oznacza chaosu, lecz fakt, że nie potrafimy deterministycznie wskazać pojedynczego wyniku, choć możemy określić rozkład możliwych rezultatów. Jeśli ten rozkład zależy od aktualnego stanu, otrzymujemy podstawową intuicję łańcucha Markowa. W najprostszym modelu własność Markowa oznacza, że prawdopodobieństwo przyszłego stanu, przy znajomości teraźniejszego, nie wymaga znajomości całej wcześniejszej historii. Ten warunek "bezpamięciowości" nie oznacza braku pamięci w potocznym sensie —

aktualny stan może przecież syntetyzować konsekwencje przeszłości. Jest to specyficzna własność warunkowej zależności.

Redukcja modelu a pluralizm metod poznawczych

Późniejsza teoria procesów stochastycznych rozwinęła tę ideę w potężny aparat matematyczny, znajdujący zastosowanie w fizyce statystycznej, ekonomii, biologii molekularnej, epidemiologii, teorii kolejek, genetyce populacyjnej, rozpoznawaniu mowy oraz uczeniu ze wzmocnieniem. Najnowsze badanie z 2026 roku, powtarzające eksperyment na Onieginie, przypomina, że historyczne źródło tej szerokiej rodziny metod tkwiło właśnie w próbie matematycznego opisu zależności w dziele literackim. Na tym poziomie nauki ścisłe i humanistyka spotykają się w sposób znacznie ciekawszy niż poprzez dekoracyjne porównanie poezji do matematyki. Markow nie tłumaczył sensu Eugeniusza Oniegina; nie rekonstruował ironii narratora, pozycji Tatiany, antropologii społecznej rosyjskiej szlachty ani relacji między Puszkinem a romantyzmem.

Dokonał on świadomej redukcji przedmiotu badania. W kontekście jego pytania badawczego utwór przestał być całością estetyczną, stając się sekwencją dwóch rodzajów elementów. Metodologicznie przypomina to praktykę modelowania typową dla całej nauki: aby badać konkretną własność układu, abstrahuje się od ogromnej liczby innych cech. Model gazu nie przechowuje zapachu powietrza, mapa meteorologiczna nie oddaje koloru każdego drzewa, a model epidemiologiczny pomija biografie zakażonych osób.

Redukcja ta nie jest błędem, o ile pamiętamy, jakie pytania umożliwia, a jakich nie może rozstrzygnąć. Tenenowska interpretacja „desakralizacji słowa” wymaga jednak akademickiego doprecyzowania. Materiał źródłowy przedstawia statystyczną analizę Puszkina jako symboliczne odarcie literatury z romantycznej aury i przejście "od hermeneutyki do statystyki". Jako figura historii intelektualnej jest to ujęcie trafne, jeśli rozumiemy je jako pojawienie się nowego reżimu poznawczego. Nie oznacza ono jednak, że probabilistyka zastąpiła hermeneutykę lub wykazała jej naukową zbędność. Markow odpowiadał na pytanie o zależności znaków, podczas gdy hermeneuta pyta o znaczenie. Obie metody mogą badać ten sam materialny tekst, lecz konstruują różne przedmioty epistemiczne. Współczesne digital humanities opierają się właśnie na takim pluralizmie: stylometria, analiza sieciowa, topic modeling i modele językowe ujawniają regularności niewidoczne podczas lektury jednostkowej, jednak ich wyniki wymagają interpretacji historycznej i kulturowej. Konflikt między „liczeniem” a „czytaniem” jest zatem źle postawiony, jeśli zakłada zastąpienie jednej techniki drugą; znacznie produktywniejsze jest pytanie o to, jak różne poziomy opisu mogą się wzajemnie ograniczać i korygować.

Biografia Markowa nadaje tej historii wymiar filozoficzno-polityczny, lecz i tutaj należy oddzielić źródłową narrację od faktów. Notatka wiąże jego probabilistykę z buntem religijnym, sugerując, że w 1912 roku wykorzystywał rachunek prawdopodobieństwa do polemiki z cudami biblijnymi. Rzeczywiście, Markow był zdecydowanym antyklerykałem i ateistą; sprzeciwiał się represjom wobec Lwa Tołstoja, a w lutym 1912 roku zwrócił się do Świętobliwego Synodu z prośbą o wykluczenie go z Cerkwi. W piśmie powoływał się m.in. na własny podręcznik rachunku prawdopodobieństwa, wyrażając sceptycyzm wobec przekazów religijnych o zdarzeniach nadprzyrodzonych.

Archiwalia wskazują jednak, że Synod nie tyle formalnie go „ekskomunikował", co uznał, iż Markow sam odłączył się od Kościoła. Związek między tą polemiką a analizą Oniegina należy zatem postrzegać jako element wspólnej intelektualnej postawy Markowa – sceptycznej wobec nieuzasadnionych założeń – a nie jako prostą tezę, że łańcuchy Markowa powstały w celu obalenia religii. Dla filozofii prawdopodobieństwa kluczowe jest coś innego: Markow przyczynił się do osłabienia intuicji, jakoby stabilność statystyczna wymagała niezależności zdarzeń. Wykazując, że odpowiednio kontrolowana zależność również pozwala uzyskać prawa graniczne, sprawił, iż probabilistyka mogła objąć znacznie szerszą klasę realnych procesów. Współczesnym językiem określilibyśmy to jako przejście od zainteresowania samą częstotliwością stanów do analizy struktury przejść. Jest to przejście od statyki rozkładu do dynamiki procesu.

Informacja jako redukcja niepewności bez znaczenia

W języku analogiczny ruch okazuje się kluczowy: nie wystarczy wiedzieć, jak często występuje litera albo słowo; trzeba wiedzieć, w jakim otoczeniu się pojawia, z czym współwystępuje i jakie kontynuacje czyni bardziej prawdopodobnymi.

Historia prowadzi nas ku Claude'owi Shannonowi, choć nie jako do prostego kontynuatora Markowa. Shannon rozwiązywał inny problem: jak matematycznie opisać komunikację przez kanał, szczególnie w obecności szumu, oraz jakie są granice kodowania i transmisji informacji. Jego praca A Mathematical Theory of Communication, opublikowana w dwóch częściach w Bell System Technical Journal w 1948 roku, stworzyła język pojęciowy, bez którego trudno wyobrazić sobie współczesną informatykę, telekomunikację i teorię kodowania. Materiał źródłowy słusznie podkreśla jeden z najbardziej radykalnych aspektów tej teorii: w obliczu technicznego problemu komunikacji Shannon strategicznie oddzielił transmisję symboli od kwestii ich znaczenia.

Nie należy jednak rozszerzać tej decyzji do filozoficznego twierdzenia, że znaczenie nie istnieje lub jest nieważne. Shannon sformułował tezę znacznie precyzyjniejszą: semantyczne aspekty

komunikatu nie są bezpośrednio istotne dla inżynieryjnego problemu wiernego wyboru i transmisji wiadomości z określonego zbioru możliwości. Ta redukcja pozwoliła zdefiniować informację poprzez niepewność. Komunikat jest informacyjny w takim zakresie, w jakim redukuje wcześniejszą niepewność odbiorcy co do tego, który z możliwych komunikatów został wybrany. Zdarzenie bardzo przewidywalne wnosi mało informacji w sensie Shannonowskim, natomiast zdarzenie nieoczekiwane wnosi jej więcej. Entropia źródła staje się zatem miarą przeciętnej niepewności rozkładu możliwych symboli.

Jest to pojęcie formalne, a nie miara mądrości, prawdy czy wartości kulturowej. Zdanie fałszywe może nieść dużo informacji Shannonowskiej, zdanie prawdziwe może nieść jej mało, a najgłębszy metafizyczny banał może mieć z punktu widzenia teorii transmisji podobny status jak dowolny inny ciąg symboli o tej samej strukturze probabilistycznej. To rozróżnienie między informacją syntaktyczno-probabilistyczną a semantyczną stało się jednym z głównych źródeł późniejszych nieporozumień w kulturze cyfrowej. Kiedy potocznie mówimy, że społeczeństwo "ma więcej informacji", mieszamy co najmniej trzy problemy: ilość przesyłanych danych, redukcję niepewności oraz wzrost wiedzy. Tymczasem wzrost przepływu danych nie gwarantuje wzrostu wiedzy, a wzrost wiedzy nie gwarantuje mądrości. Luciano Floridi i inni filozofowie informacji próbowali później przywrócić wymiar semantyczny, odróżniając informację jako odpowiednio uformowane i znaczące dane od czystej transmisji sygnału.

Shannonowska decyzja była metodologicznie niezwykle produktywna właśnie dlatego, że nie próbowała rozwiązać wszystkich problemów naraz. Dla genealogii modeli językowych szczególne znaczenie mają Shannonowskie "kolejne przybliżenia do języka angielskiego". W oryginalnym artykule konstruował on próbki tekstu o rosnącym stopniu zależności statystycznej. W przybliżeniu zerowego rzędu znaki wybierano niezależnie i jednakowo prawdopodobnie; w pierwszym rzędzie zachowywano rzeczywiste częstości liter; w kolejnym uwzględniano zależności dwu- i trzyliterowe, by następnie przejść do słów oraz prawdopodobieństw przejść między nimi. Wraz z dodawaniem zależności wygenerowane ciągi stawały się coraz bardziej podobne powierzchniowo do angielszczyzny, mimo że system nie posiadał semantycznego modelu świata. Właśnie to doświadczenie nadaje Tenenowskiej genealogii jej szczególną siłę: poprawność powierzchniowa może wyłaniać się w znacznej mierze z samej struktury statystycznej języka. Trzeba przy tym odróżnić Markowowską zależność stanów od późniejszego modelu n-gramowego. W n-gramie prawdopodobieństwo następnego elementu szacuje się na podstawie ograniczonego kontekstu poprzedzających elementów. Jeśli modelem jest bigram słowny, następne słowo zależy od słowa poprzedniego; trigram wykorzystuje dwa poprzednie słowa, a wyższe rzędy odpowiednio dłuższy kontekst.

Wraz ze wzrostem rzędu rośnie zdolność uchwycenia lokalnej struktury, lecz dramatycznie wzrasta również liczba możliwych kombinacji i problem rzadkości danych. Statystyczne NLP XX wieku wypracowało cały aparat wygładzania, estymacji i back-offu, aby radzić sobie z tymi trudnościami. Współczesne transformery rozwiązują zależność od kontekstu w radykalnie inny sposób: dzięki mechanizmowi uwagi mogą warunkować reprezentację tokenu na znacznie szerszym kontekście i nie są modelami Markowa skończonego rzędu w klasycznym sensie. Dlatego historyczna linia od Markowa do LLM jest genealogią probabilistycznego traktowania sekwencji, a nie tożsamością matematycznych mechanizmów.

Znaczenie jako statystyczny ślad ludzkich praktyk

Równolegle do teorii informacji rozwijała się lingwistyka dystrybucyjna, której epistemologia była w pewnych aspektach równie rewolucyjna. Zellig Harris w artykule Distributional Structure z 1954 roku argumentował, że strukturę języka można badać poprzez systematyczny opis względnych wystąpień jego elementów, bez konieczności wprowadzania znaczenia jako pierwotnej zmiennej analizy. John Rupert Firth sformułował natomiast w 1957 roku słynną zasadę, według której słowo poznaje się po „towarzystwie", w jakim występuje. Ta maksyma stała się jednym z pomostów łączących klasyczne językoznawstwo ze współczesnymi modelami językowymi. Należy jednak pamiętać, że Firthowska teoria kontekstu była znacznie bogatsza niż dzisiejsze hasło distributional semantics; obejmowała ona kolokację, kontekst sytuacji, funkcje językowe oraz społeczne warunki użycia. Współczesna „hipoteza dystrybucyjna" jest zatem w pewnym sensie późniejszą abstrakcją szerszego programu językoznawczego.

Zasadnicza intuicja pozostaje jednak niezwykle silna: jeśli dwa słowa systematycznie występują w podobnych kontekstach, można wnioskować, że istnieje między nimi rodzaj podobieństwa funkcjonalnego lub semantycznego. „Lekarz", „chirurg" i „pediatra" pojawiają się w innych zdaniach niż „lokomotywa", „prom" czy „śmigłowiec" – to właśnie różnice rozkładów kontekstowych tworzą statystyczny ślad różnic znaczeniowych. Nie oznacza to, że znaczenie sprowadza się wyłącznie do rozkładu słów, lecz że społecznie utrwalony język przechowuje ogromną ilość informacji o sensach właśnie poprzez strukturę użycia. Ponieważ teksty są wytwarzane przez ludzi żyjących w rzeczywistym świecie, statystyczne relacje między słowami pośrednio kodują regularności sfery społecznej, fizycznej i kulturowej.

Można tu dostrzec interesujące podobieństwo do myśli późnego Wittgensteina. W Dociekaniach filozoficznych znaczenie zostaje powiązane z użyciem słów w konkretnych praktykach językowych, a koncepcja „gry językowej" podważa poszukiwanie jednej, abstrakcyjnej esencji znaczenia

niezależnej od kontekstu działania. Nie wolno oczywiście utożsamiać Wittgensteinowskiej pragmatyki z hipotezą dystrybucyjną: użycie u Wittgensteina jest osadzone w ludzkich formach życia, normach i praktykach, podczas gdy statystyczna reprezentacja ogranicza się do tekstowego współwystępowania. Podobieństwo polega raczej na odejściu od słownikowego esencjalizmu, zgodnie z którym każde słowo posiada znaczenie niczym przedmiot ukryty za znakiem. Zarówno filozofia późnego Wittgensteina, jak i lingwistyka dystrybucyjna przesuwają punkt ciężkości ku relacjom i praktykom, choć czynią to w odmiennych reżimach metodologicznych.

Geometria znaczenia i statystyczna specyficzność terminów

W latach sześćdziesiątych i siedemdziesiątych relacyjność ta zyskała geometrię. Gerard Salton i jego współpracownicy rozwijali modele wyszukiwania informacji, w których dokumenty oraz zapytania reprezentowano jako wektory w wielowymiarowej przestrzeni terminów. Klasyczna publikacja Saltona, Wonga i Yanga z 1975 roku formalizowała vector space model dla automatycznego indeksowania.

Założenie było proste: dokument nie musi być traktowany jako literacka całość, lecz jako punkt określony zestawem współrzędnych odpowiadających ważonym terminom. Podobieństwo między dokumentami lub między dokumentem a zapytaniem można wówczas mierzyć geometrycznie – na przykład poprzez kąt pomiędzy wektorami. Znaczenie operacyjne zostaje w ten sposób przekształcone w problem odległości i kierunku w przestrzeni. Przejście od klasyfikacji do geometrii ma istotny wymiar filozoficzny. Podczas gdy Wilkins próbował umieścić pojęcia w taksonomicznych przegródkach, system przestrzeni wektorowej nie wymaga, aby każdy dokument należał do jednej rozłącznej klasy. Może on znajdować się jednocześnie blisko wielu obszarów semantycznych. Kategorie stają się stopniowalne, a podobieństwo zastępuje sztywną klasyfikację zero-jedynkową. W kognitywistyce analogiczne idee pojawiały się w modelach przestrzeni pojęciowych i reprezentacji podobieństwa. W sztucznej inteligencji przestrzeń wektorowa okazała się później idealnym medium dla reprezentacji rozproszonych, gdzie obiekt definiuje nie pojedyncza etykieta, lecz konfiguracja wielu wymiarów.

Karen Spärck Jones dodała do tej historii zasadę równie prostą, co kluczową dla wyszukiwania informacji. W artykule z 1972 roku zaproponowała statystyczną interpretację „specyficzności" terminu: słowo występujące rzadko w całej kolekcji dokumentów jest zazwyczaj bardziej rozróżniające niż słowo pojawiające się niemal wszędzie. Doprowadziło to do sformułowania idei inverse document frequency, łączonej później z częstotliwością terminu w dokumencie w rodzinie

metod TF-IDF. Jeśli dwa dokumenty zawierają słowo pospolite, wspólność ta niewiele mówi o ich tematycznym podobieństwie; jeśli jednak oba zawierają termin rzadki i charakterystyczny, zgodność jest znacznie bardziej informacyjna. Spärck Jones świadomie przesunęła „specyficzność" z poziomu słownikowego znaczenia na poziom statystycznego użycia w kolekcji.

Rozwiązanie to jest bliskim kuzynem Shannonowskiej teorii informacji. Rzadkie zdarzenie niesie więcej informacji niż zdarzenie niemal pewne; rzadki termin może być bardziej dyskryminujący niż termin powszechny. Nie są to identyczne teorie ani ten sam cel matematyczny, lecz obie korzystają z zależności między nieoczekiwanością a wartością sygnału. W ten sposób probabilistyka, telekomunikacja, bibliotekoznawstwo i językoznawstwo zaczynają tworzyć wspólne zaplecze współczesnej nauki o informacji. Historia AI nie jest tu linią biegnącą wyłącznie przez informatykę; obejmuje wkład bibliotekoznawstwa, indeksowania dokumentów, teorii komunikacji, statystyki i humanistyki języka. Z tych tradycji wyrosną później reprezentacje znacznie gęstsze niż klasyczny TF-IDF.

Latent Semantic Analysis podjęła próbę redukcji wymiaru przestrzeni dokument-termin, by wydobywać ukryte współzależności za pomocą metod algebry liniowej. Modele neuronowe z kolei zaczęły uczyć reprezentacji, w których podobne słowa otrzymują bliskie wektory nie dzięki ręcznie określonym cechom, lecz dlatego, że położenie wektora optymalizuje zdolność przewidywania kontekstu. Word2vec, GloVe i późniejsze contextual embeddings rozwinęły tę ideę jeszcze dalej: reprezentacja słowa przestała być stałym punktem niezależnym od zdania, a stała się funkcją konkretnego kontekstu. „Bank" w zdaniu o kredycie i „bank" w zdaniu o brzegu rzeki mogą otrzymać odmienne reprezentacje wewnętrzne, ponieważ nowoczesny model koduje sens zależnie od otaczających tokenów.

Taki rozwój skłania do rewizji jednej z najmocniejszych tez Tenena.

Probabilistyczna koherencja to nie epistemiczne uzasadnienie

Materiał źródłowy opisuje system statystyczny jako model, w którym „oatmeal" jest przede wszystkim punktem w przestrzeni bliskim słowom takim jak „breakfast", podczas gdy dziecko poznaje owsiankę poprzez temperaturę, smak i teksturę. To rozróżnienie pozostaje kluczowe, gdyż doświadczenie sensoryczno-motoryczne i społeczne nie jest równoważne tekstowej dystrybucji. Nie można jednak z tego wyprowadzać tezy, że reprezentacja dystrybucyjna jest całkowicie pozbawiona treści semantycznej w sensie funkcjonalnym.

W języku zapisane są bowiem rezultaty doświadczeń milionów ludzi: piszą oni o gorącu, śniadaniu, garnkach, mleku, diecie, smaku i głodzie. System analizujący gigantyczny korpus może zatem odzyskać część relacyjnej struktury pojęć bez konieczności samodzielnego jedzenia owsianki. Współczesna kognitywistyka prowadzi właśnie spór o zakres informacji konceptualnej, którą można przyswoić z samego języka, oraz o to, jaka część znaczenia wymaga bezpośredniego zakotwiczenia w percepcji i działaniu.

W tym miejscu teoria Markowa spotyka się z filozofią reprezentacji. Żaden z opisanych systemów nie przechowuje „rzeczywistości samej". Markow redukuje tekst do sekwencji dwóch stanów. Shannon sprowadza komunikat do wyboru spośród dostępnych możliwości. Salton zamienia dokument w wektor terminów, a Spärck Jones interpretuje specyficzność poprzez częstość występowania w kolekcji. Współczesny embedding reprezentuje element przez położenie w przestrzeni wyuczonej w procesie optymalizacji. Każdy z tych etapów jest redukcją, lecz każda taka redukcja umożliwia określony rodzaj obliczeń.

Epistemologicznie oznacza to, że sukces modelu nie dowodzi, iż jego reprezentacja jest ontologicznie identyczna z reprezentowanym światem. Dowodzi jedynie, że zachowuje ona pewne relacje wystarczająco dobrze dla konkretnych zadań. Taka ostrożność jest niezbędna również wtedy, gdy przechodzimy od Markowa do współczesnych LLM. Transformer nie jest „Markowem na sterydach".

Mechanizm self-attention, wielowarstwowe reprezentacje, uczenie gradientowe, tokenizacja i skala danych tworzą zupełnie inną klasę systemów. Podobieństwo historyczne polega na czym innym: zarówno Markow, jak i współczesne modele traktują obserwowany język jako źródło regularności probabilistycznych, które można poznać bez ręcznego kodowania kompletnej teorii semantycznej. W eksperymencie Markowa kontekst obejmował bardzo prostą relację stanów; w transformerze zależności mogą być rozłożone na tysiące pozycji i wiele wymiarów reprezentacji.

Następny token jest prognozowany na podstawie rozbudowanego stanu kontekstowego, a nie wyłącznie poprzedniego elementu. Z tego powodu analogia powinna wyjaśniać genealogiczne podobieństwo, a nie zacierać technicznych różnic. Jeszcze istotniejsza jest jednak różnica między przewidywaniem a prawdą. Model probabilistyczny odpowiada na pytanie, jaka kontynuacja jest prawdopodobna w świetle rozkładu, który nauczył się reprezentować. Prawdziwość wypowiedzi jest natomiast relacją pomiędzy twierdzeniem a światem, a w niektórych teoriach również praktykami uzasadniania, obserwacją i siecią innych przekonań. Wysokie prawdopodobieństwo językowe może korelować z prawdą, ponieważ teksty w dużej mierze opisują świat, ale korelacja nie stanowi logicznej gwarancji. Tutaj Tenen wraca do Ibn Chalduna: procedura może odnaleźć strukturę języka, nie uzyskując automatycznie dostępu do stanu rzeczy poza nim. Teza o „więzieniu języka"

jest najbardziej przekonująca właśnie wtedy, gdy rozumiemy ją nie jako stwierdzenie absolutnej semantycznej pustki modeli, lecz jako ostrzeżenie przed utożsamieniem probabilistycznej koherencji z epistemicznym uzasadnieniem. To rozróżnienie pozwala również lepiej zrozumieć współczesne „halucynacje" modeli generatywnych; nie należy bowiem twierdzić, że są one metafizycznie nieusuwalną właściwością każdej maszyny operującej symbolami.

Statystyczna predykcja nie jest tożsama z rozumieniem

Systemy można wzbogacać o retrieval, narzędzia, wykonanie kodu, bazy danych czy zewnętrzne procedury weryfikacji. Możliwa jest również modyfikacja celów treningowych, strategii dekodowania i mechanizmów oceny. Problem strukturalny pozostaje jednak aktualny: jeśli podstawowym zadaniem systemu jest przewidywanie kolejnych elementów języka, sama wysoka jakość tej predykcji nie ustanawia relacji dowodowej między wygenerowanym twierdzeniem a rzeczywistością. Tutaj genealogia Markow-Shannon-Firth pomaga wyjaśnić zarówno siłę, jak i granice współczesnych modeli. Statystyczna struktura języka zawiera więcej wiedzy, niż przypuszczały wcześniejsze pokolenia informatyków, lecz nie dostarcza automatycznie wszystkich procedur niezbędnych do prawdziwego uzasadniania twierdzeń.

Znaczenie tej historii dla filozofii inteligencji jest szersze. Tradycyjny racjonalizm skłaniał do poszukiwania wiedzy w jawnych regułach, definicjach i dedukcjach. Podejście probabilistyczne pokazuje natomiast, że część kompetencji może polegać na uchwyceniu regularności bez możliwości wyrażenia ich w postaci prostego zbioru zasad. Człowiek również często przewiduje, rozpoznaje i klasyfikuje, nie potrafiąc sformułować wszystkich przesłanek swojego działania.

Psychologia poznawcza od dawna bada uczenie implicytne, heurystyki, rozpoznawanie wzorców oraz statystyczne uczenie się u dzieci. Neuronauka analizuje mózg jako system przewidywania i aktualizacji reprezentacji pod wpływem błędu predykcji. Koncepcje takie jak Bayesian brain, predictive coding i predictive processing nie są tożsame z modelem językowym, ale dowodzą, że probabilistyczność sama w sobie nie stanowi argumentu przeciw poznaniu. Spór powinien zatem dotyczyć rodzaju reprezentacji, źródeł danych, zdolności do działania, ucieleśnienia, generalizacji i korekty błędu, a nie samego faktu wykorzystywania prawdopodobieństwa.

Markow odwrócił jeden z najstarszych filozoficznych stereotypów: przypadek i rozum przestały być przeciwieństwami. Proces probabilistyczny może posiadać strukturę, niepewność może stać się przedmiotem ścisłej matematyki, a przewidywanie może być racjonalne nawet bez deterministycznej pewności. Dla epistemologii jest to wniosek kluczowy. Wiedza empiryczna rzadko przyjmuje postać absolutnej konieczności; nauki przyrodnicze i społeczne operują

rozkładami, przedziałami ufności, prawdopodobieństwami warunkowymi, ryzykiem oraz aktualizacją przekonań. Dzisiejsza AI jest więc dziedzicem nie tylko historii maszyn, lecz także wielkiej zmiany kulturowej, w której nowoczesna nauka nauczyła się traktować niepewność jako wielkość możliwą do formalnego analizowania.

Zarazem statystyczny sukces generuje własne ryzyko filozoficzne. Gdy system coraz lepiej przewiduje tekst, łatwo przejść od tezy instrumentalnej („model przewiduje regularności językowe") do tezy ontologicznej („model posiada taki sam rodzaj rozumienia jak człowiek"). Błąd odwrotny jest równie prosty: skoro model nie doświadcza świata biologicznie, jego reprezentacje miałyby być całkowicie pozbawione znaczenia. Obie tezy są zbyt radykalne.

Interdyscyplinarna analiza wymaga rozdzielenia kompetencji behawioralnej, reprezentacji statystycznej, semantycznego odniesienia, świadomości fenomenalnej, sprawczości praktycznej i odpowiedzialności moralnej. To, że system wykazuje biegłość w jednym z tych wymiarów, nie przesądza automatycznie o pozostałych.

Tenenowska genealogia osiąga dzięki Markowowi punkt, w którym wcześniejsza „magia liter" zostaje całkowicie odczarowana matematycznie, lecz paradoksalnie odzyskuje część swojej dawnej mocy. Zairadża obiecywała odkrywać nieznane poprzez transformację znaków, nie posiadając naukowej teorii działania tej procedury. Markow wskazał realny mechanizm, dzięki któremu jeden stan symboliczny rzeczywiście zawiera informację probabilistyczną o następnym. Shannon zmierzył tę informację i uczynił z niej fundament inżynierii komunikacji. Harris i Firth wykazali, że kontekst dystrybucyjny może być źródłem wiedzy o strukturze języka, natomiast Salton i Spärck Jones przekształcili tekst w geometrię i statystyczną wagę.

Kolejne generacje modeli neuronowych uczą się coraz bogatszych przestrzeni reprezentacji. To, co w średniowiecznej kombinatoryce było metafizyczną nadzieją na wydobycie wiedzy ze znaków, w XX wieku stało się empirycznie sprawdzalnym programem naukowym. Nie oznacza to jednak, że sam język wystarcza do zbudowania kompletnej teorii inteligencji.

Wręcz przeciwnie – im większy sukces osiągają modele statystyczne, tym pilniejsze staje się pytanie o to, czego ten sukces właściwie dowodzi. Jeżeli tekst pozwala odtworzyć ogromną część ludzkiej wiedzy pojęciowej, należy wyjaśnić, jak wiele informacji o świecie zostało wcześniej zakodowanych w społecznym użyciu języka.

Od statystyki znaków do wielowymiarowych reprezentacji kontekstowych

Jeżeli modele zawodzą w zadaniach wymagających niezawodnego kontaktu z rzeczywistością, należy ustalić, czy przyczyną jest brak odpowiednich danych, ucieleśnienia, ograniczenia architektury, sposób optymalizacji czy brak zewnętrznego sprzężenia zwrotnego. Gdy system potrafi rozwiązywać nowe problemy, trzeba rozróżnić interpolację w ogromnej przestrzeni reprezentacji od generalizacji wymagającej abstrakcyjnej rekonstrukcji reguły. Spór o AI po Markowie nie powinien więc sprowadzać się do pytania, czy statystyka „jest" rozumieniem, lecz do badania, jakie klasy struktur poznawczych mogą wyłaniać się z probabilistycznego uczenia i jakie dodatkowe warunki są niezbędne do uzyskania kompetencji przekraczających tekst. Przykład „Pushkina Markowa" prowadzi nas wprost do centralnego problemu współczesnej sztucznej inteligencji.

Wszystko zaczęło się od samogłosek i spółgłosek, ale nie kończy na prognozowaniu znaków. Kolejnym etapem było przejście od lokalnych zależności sekwencyjnych do wielowymiarowych reprezentacji kontekstowych; od ręcznie projektowanych cech do uczenia reprezentacji, od wyszukiwania dokumentów do generowania języka oraz od modeli statystycznych opisujących istniejący tekst do systemów zdolnych wytwarzać nową treść w odpowiedzi na instrukcję. Dopiero na tym tle można uczciwie postawić pytanie, które powraca od Kirchera i Kuhlmanna, lecz w XXI wieku nabiera zupełnie nowej skali: jeżeli z probabilistycznego przewidywania wyłania się zachowanie funkcjonalnie przypominające rozumowanie, to jakiego dokładnie rodzaju jest ta własność, gdzie leżą jej granice i dlaczego ludzki obserwator tak chętnie nazywa ją „inteligencją"?

Genealogia prowadząca od Markowowskiego „Eugeniusza Oniegina" ku współczesnym dużym modelom językowym nie jest prostą historią zwiększania liczby parametrów ani technicznego doskonalenia predykcji kolejnego słowa. Jej istotą jest stopniowa ewolucja pojęcia reprezentacji. W modelu Markowa znaczenie konkretnej litery nie było przedmiotem analizy; kluczowe stawały się prawdopodobieństwa przejścia między stanami. W klasycznym wyszukiwaniu informacji dokument przedstawiano jako punkt w przestrzeni terminów, natomiast w reprezentacjach dystrybucyjnych położenie słowa zaczęło kodować podobieństwa wynikające ze struktury jego użycia. W sieciach neuronowych reprezentacja przestała być projektowana bezpośrednio przez badacza – zaczęła być wyuczana jako wewnętrzny stan umożliwiający skuteczne rozwiązanie zadania optymalizacyjnego.

Transformer dopełnił ten proces, wprowadzając możliwość dynamicznego rekonstruowania znaczenia jednostki w zależności od szerokiego kontekstu. Z perspektywy filozofii poznania zmiana ta jest zasadnicza: współczesny model językowy nie posiada słownika pojęć w sensie klasycznej

ontologii symbolicznej. Dysponuje natomiast wielowymiarową architekturą relacji, w której znaczenie operacyjne wyłania się jako własność struktury reprezentacyjnej.

Od statystycznych wektorów do kontekstowych reprezentacji języka

Tenen interpretuje tę ewolucję jako przedłużenie znacznie starszej historii „literackich maszyn": od kombinatoryki i klasyfikacji, przez statystyczne sąsiedztwo, aż po modele generujące język bez uprzednio zdefiniowanej, ręcznie zapisanej teorii świata. Autor wskazuje na Firthowską zasadę poznawania słowa poprzez jego „towarzystwo", rozwój przestrzeni wektorowych oraz statystyczne ważenie terminów jako bezpośrednie intelektualne zaplecze współczesnych LLM. Jednocześnie Tenen zachowuje wobec tego sukcesu zasadniczą rezerwę: bogactwo mapy relacji językowych nie pozwala, by bez dodatkowej argumentacji utożsamiać jej z biologicznie i społecznie ucieleśnionym poznaniem człowieka. W jego szerszym ujęciu „metafora nie jest mechanizmem" – podobieństwo zachowania modelu do ludzkiego nie przesądza o identyczności procesów, które do tego zachowania prowadzą.

Pierwszym istotnym krokiem od klasycznej przestrzeni dokumentów ku współczesnym reprezentacjom neuronowym były gęste wektory słów. Modele word2vec Tomasza Mikolova i jego współpracowników z 2013 roku wykazały, że możliwe jest efektywne uczenie ciągłych reprezentacji na ogromnych korpusach, a geometria uzyskanej przestrzeni odzwierciedla zarówno regularności syntaktyczne, jak i semantyczne. Rok później Jeffrey Pennington, Richard Socher i Christopher Manning przedstawili GloVe, łączący globalne współwystępowanie słów z geometrycznym modelem wektorowym. Wykazali oni, że relacje znaczeniowe mogą ujawniać się jako regularności kierunków i różnic między wektorami. Nie należy jednak wyciągać naiwnego wniosku, że semantyka została „rozwiązana" przez geometrię. Kluczowy był dowód, że znaczna część struktury językowej – którą wcześniejsze systemy próbowały opisać za pomocą jawnych reguł i słowników – może zostać odtworzona automatycznie z rozkładów użycia. Filozoficznie był to triumf reprezentacji relacyjnej nad reprezentacją esencjalistyczną.

W tradycyjnym słowniku pojęcie otrzymuje definicję określającą jego zakres i treść. W embeddingu słowo zyskuje współrzędne, których znaczenie ujawnia się przede wszystkim w relacjach do innych punktów przestrzeni. Nie istnieje pojedynczy wymiar zatytułowany „królewskość", „kobiecość" czy „przeszłość"; właściwości te są zakodowane rozproszone w konfiguracji wielu wymiarów. Takie reprezentacje stanowiły oś sporu między symboliczną sztuczną inteligencją a koneksjonizmem. Symboliczna AI, z Newellem i Simonem na czele, zakładała, że procesy inteligentne można opisać

jako manipulowanie dyskretnymi strukturami reprezentującymi obiekty i relacje. Koneksjonizm szukał natomiast wiedzy w rozkładzie aktywacji i wag sieci.

Współczesne systemy generatywne nie rozstrzygają tego sporu w sposób jednoznaczny: choć reprezentacje neuronowe są rozproszone, ich działanie prowadzi do produkcji dyskretnych symboli, a praktyczne rozwiązania coraz częściej łączą modele neuronowe z narzędziami symbolicznymi, kodem i bazami danych. Word2vec czy GloVe posiadały zasadnicze ograniczenie: pojedynczy wyraz otrzymywał jedną reprezentację, niezależnie od zdania. Tymczasem język naturalny jest radykalnie kontekstowy. Znaczenie „banku" w kontekście kredytu różni się od znaczenia brzegu rzeki, a ironia czy metafory mogą przesuwać funkcję słowa jeszcze dalej. Modele kontekstowe zmieniły ontologię embeddingu: jednostka językowa nie musi posiadać jednej stałej reprezentacji, gdyż może być ona wyliczana na nowo w zależności od otoczenia.

Architektura transformerów jako matematyczna technologia dynamicznego kontekstu

Ostatecznie właśnie ta zasada stała się podstawą działania transformerów. Artykuł Attention Is All You Need z 2017 roku zaproponował architekturę, która zrezygnowała z sekwencyjnego przetwarzania informacji, charakterystycznego dla wcześniejszych sieci rekurencyjnych, i oparła modelowanie zależności przede wszystkim na mechanizmie attention. Transformer umożliwił każdemu elementowi reprezentacji uwzględnianie innych części sekwencji poprzez obliczanie wag określających ich aktualną relewancję. Rozwiązanie to okazało się nie tylko skuteczne jakościowo, ale i znacznie lepiej nadawało się do równoległych obliczeń, co miało kluczowe znaczenie dla późniejszego skalowania modeli.

Mechanizm attention można interpretować jako radykalizację Firthowskiej zasady kontekstu: element nie posiada znaczenia wyłącznie „sam w sobie", lecz jego bieżąca reprezentacja powstaje w dynamicznej relacji do innych elementów. Nie należy oczywiście utożsamiać matematycznego mechanizmu uwagi z ludzką uwagą psychologiczną; nazwa ta pełni funkcję metafory, a nie twierdzenia o tożsamości procesów. Ma to interesujące konsekwencje dla filozofii reprezentacji. W klasycznym modelu języka słownik można wyobrażać sobie jako magazyn stabilnych jednostek, z których buduje się zdania. Transformer odwraca tę intuicję: informacja o jednostce jest rekontekstualizowana na kolejnych warstwach przetwarzania. Znaczenie operacyjne staje się procesem, a nie statyczną własnością.

Techniczna architektura zaczyna tu przypominać nurty filozofii języka podkreślające kontekstualność i relacyjność znaczenia, choć podobieństwo to nie oznacza zależności genealogicznej. Gadamerowska hermeneutyka, Wittgensteinowskie znaczenie poprzez użycie czy pragmatyczne teorie języka wskazywały, że wypowiedzi nie są interpretowane przez mechaniczne odczytywanie izolowanych definicji, lecz wewnątrz szerszego horyzontu praktyk i relacji. Transformer dostarcza matematycznej technologii dynamicznego kontekstu, jednak jego „kontekst" jest obliczeniową strukturą danych, a nie ludzką formą życia.

Te dwa poziomy należy porównywać analitycznie, nie utożsamiać ontologicznie. Sama architektura nie wyjaśnia jednak skali późniejszej zmiany. Drugim istotnym elementem stało się samonadzorowane uczenie na ogromnych korpusach. W autoregresyjnym modelu językowym podstawowym zadaniem jest przewidywanie kolejnego tokenu na podstawie wcześniejszego kontekstu. To pozornie skromne kryterium wymusza jednak przyswojenie ogromnej liczby regularności; trafna predykcja wymaga bowiem rozpoznawania składni, fleksji, stylu, zależności semantycznych, gatunków, struktur dyskursu, faktów oraz wielu pragmatycznych wzorów interakcji. Jeśli w tekście pojawia się opis reakcji chemicznej, programu komputerowego czy argumentu prawnego, prawidłowa kontynuacja może wymagać reprezentowania struktury danej domeny. Z perspektywy uczenia maszynowego jest to postać representation learning: model nie otrzymuje jawnego katalogu cech, lecz wytwarza wewnętrzne reprezentacje użyteczne dla minimalizacji błędu predykcji. Ta właściwość pozwala lepiej zrozumieć, dlaczego analogia "statystycznej papugi" jest zarazem cenna i niewystarczająca.

Model rzeczywiście nie nabywa języka jak dziecko i nie posiada biologicznej historii socjalizacji. Jednak trafne przewidywanie języka na wielką skalę może wymagać stworzenia wewnętrznych zmiennych zachowujących znacznie bogatsze zależności niż powierzchniowa częstość słów. Termin „statystyka" nie oznacza bowiem automatycznie płytkości. Mechanika statystyczna opisuje własności materii, genetyka populacyjna wykorzystuje modele probabilistyczne do badania ewolucji, a neurobiologia interpretuje zachowanie organizmu poprzez probabilistyczne kodowanie i przewidywanie. Pytanie naukowe nie brzmi zatem, czy LLM jest „tylko statystyczny", gdyż probabilistyczność sama w sobie niewiele mówi o głębokości reprezentacji. Należy raczej ustalać, jakie własności zostały zakodowane, czy są wykorzystywane przyczynowo przez model, jak generalizują oraz w jakim zakresie pozostają zależne od powierzchniowych korelacji.

Trzecim elementem rewolucji było skalowanie. Prace nad empirycznymi scaling laws wykazały, że strata predykcyjna modeli językowych zmienia się regularnie wraz z wielkością modelu, objętością danych i ilością obliczeń przeznaczonych na trening. Badania Kaplana i współpracowników z 2020 roku wskazały na istnienie przybliżonych praw potęgowych w szerokim zakresie skal, co

sugerowało, że postęp można planować poprzez zwiększanie odpowiednio dobranych zasobów zamiast każdorazowego oczekiwania na nową architekturę. Późniejsza praca Hoffmanna i współautorów nad modelem Chinchilla zmodyfikowała to rozumienie: przy ograniczonym budżecie obliczeniowym nie należy zwiększać liczby parametrów w oderwaniu od ilości danych, lecz skalować oba składniki bardziej równomiernie. Model o 70 miliardach parametrów, trenowany na większej liczbie tokenów, potrafił przewyższać znacznie większe modele trenowane mniej optymalnie. W ekonomii technologii oznacza to, że „inteligencja modelu" nie jest funkcją jednej magicznej zmiennej, lecz rezultatem alokacji komplementarnych zasobów: architektury, danych, mocy obliczeniowej, metod optymalizacji i późniejszego dostrajania. Skalowanie zmieniło również naturę interfejsu między modelem a użytkownikiem.

Reprezentacje wewnętrzne nie są dowodem na istnienie umysłu

GPT-3 zaprezentował w 2020 roku znaczącą poprawę w zakresie tzw. few-shot learning: model zyskał zdolność wykonywania nowych zadań na podstawie instrukcji lub kilku przykładów podanych bezpośrednio w kontekście, bez konieczności klasycznego dostrajania parametrów. Zjawisko to, nazwane później in-context learning, jest filozoficznie istotne, gdyż wymusza rozróżnienie między uczeniem rozumianym jako trwała zmiana wag a adaptacją zachowania na podstawie chwilowego kontekstu. Model w trakcie pojedynczej rozmowy nie musi modyfikować swoich parametrów, by wykorzystać przykład jako tymczasową instrukcję regulującą generację. W kognitywistyce analogia do ludzkiej pamięci roboczej lub szybkiego uczenia jest kusząca, jednak wymaga ostrożności: podobieństwo funkcjonalne nie dowodzi identyczności reprezentacji ani mechanizmu neurobiologicznego.

Jeszcze większą powściągliwość należy zachować wobec pojęcia „emergencji". W 2022 roku Wei i współpracownicy opisali pewne zdolności jako emergentne, jeśli nie były one widoczne w mniejszych modelach, a pojawiały się w większych, wymykając się prostej ekstrapolacji wcześniejszych wyników. Termin ten szybko wykroczył poza precyzyjny sens techniczny, wiążąc się w kulturze popularnej z narracją o nieprzewidywalnym „budzeniu się" inteligencji. Późniejsza analiza Schaeffera, Mirandy i Koyejo wykazała jednak, że część pozornie nagłych progów wynika ze stosowanych metryk. Przy użyciu funkcji oceny ciągłych zamiast progowych lub opartych na dokładnym dopasowaniu, wzrost kompetencji jawi się jako proces znacznie bardziej stopniowy. Współczesne przeglądy wciąż traktują emergencję jako otwarty problem, wymagający precyzyjniejszych definicji i oddzielenia zjawisk wynikających ze skalowania od tych związanych ze

złożonością zadania, metryką, treningiem czy sposobem promptowania. Z naukowego punktu widzenia nie ma więc podstaw, by każde nowe zachowanie większego modelu interpretować jako jakościowy skok ontologiczny od "statystyki" do "umysłu". Znacznie konkretniejsze argumenty dostarczają badania nad reprezentacjami wewnętrznymi.

Eksperyment z Othello-GPT jest o tyle interesujący, że model trenowano na przewidywaniu legalnych sekwencji ruchów w grze, nie przekazując mu jawnej reprezentacji planszy. Analiza ujawniła jednak istnienie wewnętrznego, nieliniowego stanu odpowiadającego położeniu pionów, a eksperymenty interwencyjne wykazały, że zmiana tej reprezentacji przyczynowo wpływa na dalsze przewidywania.

Wynik ten osłabia skrajną tezę, jakoby predyktor sekwencji operował wyłącznie na powierzchniowych statystykach. Aby skutecznie przewidywać, model może wytworzyć reprezentację ukrytego stanu generującego obserwowane dane. Nie oznacza to jeszcze budowy kompletnej ludzkiej ontologii świata, lecz pokazuje, że cel predykcyjny może prowadzić do powstania struktur reprezentujących przyczynowo istotne cechy środowiska. Podobnie badania nad reprezentacją przestrzeni i czasu w rodzinach modeli językowych wykazały możliwość liniowego odczytywania z aktywacji informacji odpowiadających położeniom geograficznym i relacjom temporalnym. Autorzy interpretowali to jako dowód istnienia podstawowych składników modelu świata, zaznaczając jednocześnie, że spójna reprezentacja współrzędnych nie jest jeszcze dynamicznym, przyczynowym modelem rzeczywistości. Badania geometrii reprezentacji idą dalej, próbując formalizować hipotezę, że wysokopoziomowe cechy są kodowane jako kierunki lub inne regularne struktury w przestrzeni aktywacji.

Praca Parka, Choe i Veitcha z ICML 2024 nadała „linear representation hypothesis” bardziej ścisłą postać, wiążąc ją z probingiem oraz sterowaniem zachowaniem modeli. Nowsze badania topologii reprezentacji z 2026 roku wskazują ponadto, że kolejne warstwy modeli mogą przechodzić przez różne fazy organizacji geometrycznej, co sugeruje dynamiczne przekształcanie przestrzeni reprezentacyjnej w trakcie pojedynczego przejścia sieci. Z takimi badaniami wiąże się jednak poważny problem metodologiczny.

Fakt, że z aktywacji modelu można odczytać pewną informację za pomocą klasyfikatora lub regresji, nie zawsze oznacza, że sam model wykorzystuje ją w procesie decyzyjnym. Reprezentacja może zawierać korelat dostępny dla zewnętrznego badacza, który nie steruje funkcjonalnie generacją. Dlatego kluczowe stają się eksperymenty przyczynowe: nie tylko wykrywanie informacji, lecz manipulowanie stanem wewnętrznym i sprawdzanie, czy zachowanie modelu zmienia się w przewidywany sposób. Badania Othello należą właśnie do tej silniejszej kategorii. Podobny problem dotyczy eksperymentów nad „geometrią prawdy”. Niektóre prace wykazały liniowe kierunki

związane z prawdziwością prostych twierdzeń oraz możliwość częściowego sterowania odpowiedziami przez modyfikację aktywacji, lecz nie oznacza to, że model posiada jednolitą, filozoficznie określoną kategorię prawdy. Badania z 2025 roku pokazują zresztą, że stabilność takich „truth directions" zależy od modelu i rodzaju zadania.

W tym miejscu należy ponownie skonfrontować współczesną wiedzę z mocnym sformułowaniem obecnym w materiale Tenena, według którego model pozostaje zamknięty w "więzieniu języka", a jego semantyka jest przede wszystkim geometrią sąsiedztwa. Najnowsze badania nie unieważniają tej krytyki, lecz wymagają jej doprecyzowania. W 2025 roku w Nature Human Behaviour porównano reprezentacje około 4 442 pojęć uzyskane od ludzi z tymi generowanymi przez kilka rodzin LLM.

Modele pozbawione bezpośredniego zakotwiczenia sensorycznego stosunkowo dobrze odtwarzały wiele niesensomotorycznych wymiarów ludzkich pojęć, natomiast podobieństwo systematycznie malało w domenach sensorycznych i było najniższe dla aspektów motorycznych. Dodanie doświadczenia wizualnego zwiększało zgodność na wymiarach związanych z widzeniem.

AI jako skondensowany rezultat rozproszonego systemu społeczno-technicznego

Wynik ten jest szczególnie istotny, gdyż nie potwierdza ani naiwnego przekonania, że język wystarcza do wszystkiego, ani tezy, iż z samego języka nie można wywieść żadnego rzeczywistego znaczenia. Obnaża raczej pewną asymetrię: społeczny korpus językowy przechowuje niezwykle bogate informacje konceptualne, jednak część struktury doświadczenia pozostaje trudna do odzyskania bez kontaktu z innymi modalnościami. Można to powiązać z teoriami grounded cognition, embodied oraz nurtami 4E. W tym ujęciu poznanie jest embodied, embedded, enacted i extended: cielesne, osadzone w środowisku, realizowane poprzez działanie i częściowo rozciągnięte na narzędzia. Z perspektywy radykalnego enaktywizmu znaczenie nie jest przede wszystkim reprezentacją ukrytą w głowie, lecz wyłania się z historii interakcji organizmu ze światem. Taki punkt widzenia stanowi silny kontrargument wobec tezy, że nawet niezwykle bogata reprezentacja tekstowa mogłaby stanowić wystarczającą teorię inteligencji.

Z drugiej strony wyniki współczesnych LLM dowodzą, że język sam w sobie jest produktem ucieleśnionych i społecznych organizmów. Korpus nie stanowi zbioru znaków powstałych w próżni; jest zapisem tego, co ludzie widzieli, zrobili, odczuli, zmierzyli i przekazali. Model może zatem pośrednio odzyskiwać strukturę świata z językowych śladów doświadczenia, choć droga ta jest

znacznie mniej bezpośrednia niż w przypadku organizmu uczącego się poprzez percepcję i działanie. Fakt ten wzmacnia jedną z głównych tez Tenena, lecz zmienia jej sens.

Jeśli LLM nabywa część „wiedzy o świecie” poprzez korelacje językowe, nie oznacza to, że maszyna niezależnie odkryła świat. Znaczna część tej struktury została wcześniej wprowadzona do korpusu przez ludzkie procesy poznawcze. Materiał źródłowy definiuje AI jako efekt pracy zbiorowej i przeciwstawia obrazowi autonomicznego umysłu koncepcję "mówiącej biblioteki", "personifikacji chóru" oraz "shadow team".

Jest to interpretacja o doniosłym znaczeniu socjologicznym. Reprezentacja wyuczona przez model jest technicznym osiągnięciem konkretnego systemu, jednak warunki umożliwiające jej powstanie obejmują języki rozwijane przez stulecia, praktyki piśmiennicze, instytucje edukacyjne, literaturę, dokumentację naukową oraz infrastrukturę sieciową. To także efekt pracy autorów, redaktorów, tłumaczy, bibliotekarzy, programistów, badaczy, operatorów centrów danych oraz osób przygotowujących i oceniających dane. Kompetencja ujawniająca się w jednym interfejsie jest z tej perspektywy skondensowanym rezultatem rozproszonego systemu społeczno-technicznego.

Wymusza to rozróżnienie co najmniej trzech poziomów „wiedzy” modelu. Pierwszy ma charakter statystyczny: system koduje regularności umożliwiające przewidywanie. Drugi jest funkcjonalny: reprezentacje pozwalają na wykonywanie operacji, które interpretujemy jako klasyfikację, wnioskowanie, translację, syntezę czy planowanie.

Rozpad pojęcia inteligencji na problemy epistemologiczne i instytucjonalne

Trzeci problem jest epistemologiczny: pytamy, czy system posiada warunki pozwalające określić jego odpowiedzi jako uzasadnioną wiedzę, a nie tylko trafne generacje. W filozofii wiedzy sama trafność jest niewystarczająca; przypadkowo prawdziwe przekonanie zazwyczaj nie jest traktowane jako wiedza. Analogicznie model może wygenerować prawdziwe twierdzenie bez niezawodnego mechanizmu uzasadnienia lub dysponować wewnętrzną reprezentacją trafnych relacji, a mimo to w konkretnym przebiegu wygenerować fałsz. Z tego powodu zdolność modelu do kodowania prawdy i jego zdolność do zawsze prawdziwego odpowiadania są odmiennymi problemami badawczymi.

W konsekwencji słowo „rozumowanie” również wymaga dekompozycji. W psychologii poznawczej może oznaczać operacje inferencyjne wykonywane na reprezentacjach; w logice formalnej chodzi o zachowanie relacji wynikania; w filozofii umysłu problemem jest mechanizm tworzenia i przekształcania treści, natomiast w informatyce termin ten bywa używany operacyjnie dla klasy

zadań wymagających wieloetapowej transformacji informacji. Model może wykazywać wysoką skuteczność w zadaniach nazywanych reasoning benchmarks, ale sam wynik nie przesądza, czy jego proces odpowiada ludzkiemu rozumowaniu symbolicznemu. Z drugiej strony nie można uznać, że skoro proces różni się od ludzkiego, to nie zachodzi żadna funkcjonalnie istotna inferencja. Nauka potrzebuje tutaj taksonomii mechanizmów, nie wojny metafor. Równie ostrożnego podejścia wymaga Tenenowska teza o przejściu od „general intelligence" ku „generic intelligence". Materiał źródłowy interpretuje uśredniony język modeli jako ryzyko homogenizacji ekspresji: system wytrenowany na ogromnym przekroju tekstów może preferować rozwiązania typowe, bezpieczne i statystycznie centralne. Zagrożenie jest realne, lecz nie wynika deterministycznie z samego modelu probabilistycznego.

Rozkład generacji może być bowiem modulowany kontekstem, instrukcją, próbkowaniem, dostrajaniem, narzędziami i wyspecjalizowanymi zbiorami wiedzy. Ten sam system może produkować tekst uśredniony albo bardzo specyficzny, zależnie od konfiguracji interakcji. Bardziej trafne jest więc potraktowanie „generyczności" jako własności określonego reżimu użycia i optymalizacji, a nie koniecznej istoty LLM. Problem społeczny zaczyna się wtedy, gdy ogromna liczba instytucji posługuje się podobnymi modelami, kryteriami bezpieczeństwa, systemami rankingowymi i wzorcami poprawności. Homogenizacja może wówczas powstać na poziomie infrastruktury kulturowej, nawet jeśli pojedynczy model technicznie posiada szeroką przestrzeń możliwych odpowiedzi. W tym miejscu techniczna historia modeli językowych przechodzi w socjologię instytucji, ekonomię polityczną i filozofię odpowiedzialności.

Transformer sam w sobie nie wyjaśnia, kto wybiera dane treningowe, jakie formy pracy zostają w nich niewidzialnie skumulowane, kto ponosi koszt budowy infrastruktury, kto posiada model, według jakich wartości następuje jego dostrajanie, komu przypisywane są błędy i kto przejmuje ekonomiczną wartość automatyzowanej kompetencji. Materiał Tenena konsekwentnie przesuwa interpretację inteligencji właśnie w tę stronę: AI nie powinna być analizowana jako samotny syntetyczny mózg, lecz jako instytucjonalnie i historycznie skonstruowany układ pracy zbiorowej, danych, techniki oraz decyzji organizacyjnych. Techniczna kompetencja modeli jest więc niezbędnym, ale niewystarczającym przedmiotem analizy. Im większa staje się ich funkcjonalna samodzielność, tym ważniejsze jest ustalenie, gdzie w całym systemie społeczno-technicznym znajduje się odpowiedzialność, sprawstwo i władza.

Współczesne LLM nie zamykają wielowiekowego sporu rozpoczętego przez maszyny literowe, szafy Kirchera, kombinatorykę Llulla, programowalność Babbage'a i statystykę Markowa. Przeciwnie – nadają temu sporowi znacznie większą ostrość.

Wiemy dziś, że probabilistyczne uczenie na języku może wytworzyć reprezentacje o niezwykłej złożoności, zdolne uchwycić część struktury semantycznej, przestrzennej, temporalnej i konceptualnej świata. Wiemy zarazem, że reprezentacje te nie są tożsame z ludzkim doświadczeniem sensorimotorycznym, że ich pojawienie się nie rozstrzyga problemu świadomości ani moralnego podmiotu, że wysoka trafność predykcji nie gwarantuje prawdy oraz że zachowanie systemu jest rezultatem infrastruktury pracy i instytucji niewidocznych w pojedynczym oknie dialogowym. Najważniejszym rezultatem obecnego etapu badań nie jest więc dowód na to, że „maszyna naprawdę myśli", ani dowód przeciwny. Jest nim rozpad zbyt prostego pojęcia inteligencji na kilka empirycznie i filozoficznie odmiennych problemów: reprezentację, uczenie, generalizację, planowanie, zakotwiczenie, świadomość, sprawczość, odpowiedzialność i uczestnictwo w systemie społecznym.

Genealogia techniczna może zatem ustąpić analizie politycznej i instytucjonalnej. Jeżeli bowiem model jest zdolny do wykonywania części pracy wcześniej utożsamianej z kompetencją lekarza, programisty, tłumacza, analityka, nauczyciela czy autora, zasadniczym pytaniem nie jest już wyłącznie to, co znajduje się „wewnątrz" jego reprezentacji.

Równie ważne staje się pytanie, jaka reorganizacja stosunków społecznych zachodzi wokół niego: kto dostarczył pracę skumulowaną w danych, komu przysługuje autorstwo, kto odpowiada za decyzję wspomaganą algorytmicznie, czy infrastruktura techniczna decentralizuje wiedzę czy koncentruje władzę, oraz czy automatyzacja zwiększa autonomię człowieka, czy przekształca go w operatora gotowych kompetencji.

AI jako skondensowana praca zbiorowa

To właśnie na tym poziomie dziewięć końcowych tez Tenena przestaje być zestawem efektownych maksym, a staje się programem badawczym obejmującym socjologię pracy, teorię organizacji, prawo, ekonomię, etykę i filozofię polityczną. Polityczna ekonomia sztucznego intelektu: praca zbiorowa, własność, firma i odpowiedzialność w epoce automatyzacji wiedzy.

Przejście od analizy reprezentacji wewnętrznych modeli językowych do politycznej ekonomii sztucznej inteligencji wymaga zmiany jednostki analizy. Jeśli pytamy wyłącznie o parametry, mechanizm uwagi, zdolność generalizacji czy prawdopodobieństwo kolejnego tokenu, badamy model jako artefakt techniczny. Jeśli jednak zapytamy, skąd pochodzą dane, kto wykonał pracę umożliwiającą ich powstanie, kto finansuje obliczenia i posiada infrastrukturę, kto organizuje wdrożenie systemu, kto czerpie z niego rentę ekonomiczną oraz kto ponosi odpowiedzialność za

szkodliwe konsekwencje — przedmiotem badania staje się system społeczno-techniczny. To właśnie na tym poziomie Dennis Yi Tenen formułuje swoją tezę o collective labor.

W materiale źródłowym przeciwstawia on obrazowi AI jako autonomicznego „ducha" wizję skumulowanej pracy inżynierów, leksykografów, badaczy, twórców korpusów i użytkowników, której efekty są personifikowane przez jednolity interfejs modelu. W takim ujęciu pytanie o to, czy model jest inteligentny, staje się wtórne wobec pytania o architekturę produkcji inteligencji: jakie zasoby społeczne zostały skondensowane w artefakcie, jak przekształcono je w kapitał techniczny i jakie stosunki własności organizują późniejszy dostęp do rezultatów tej kondensacji.

Tenenowskie pojęcie pracy zbiorowej można osadzić w długiej tradycji ekonomii politycznej, a szczególnie w marksowskim rozróżnieniu między pracą żywą a pracą uprzedmiotowioną. Maszyna nie powstaje jako czysta materializacja abstrakcyjnej „technologii"; jest rezultatem wcześniejszej pracy projektowej, naukowej, inżynieryjnej, organizacyjnej i produkcyjnej. Po skonstruowaniu może jednak wystąpić wobec aktualnego pracownika jako autonomicznie wyglądająca siła produkcyjna. Marks w Grundrisse rozwijał pojęcie general intellect — społecznej wiedzy, która w rozwiniętej produkcji zostaje wcielona w maszyny i organizację procesu wytwórczego. Nie należy przy tym anachronicznie twierdzić, że opisywał on w ten sposób modele językowe.

Analogia strukturalna jest jednak wyraźna: wiedza wytworzona społecznie może zostać przekształcona w kapitał trwały lub infrastrukturę techniczną, która następnie pozwala pojedynczemu właścicielowi organizować produkcję na skalę przekraczającą kompetencje każdego indywidualnego pracownika. W generatywnej AI proces ten przyjmuje szczególną postać, ponieważ uprzedmiotowieniu podlega nie tylko sprawność mechaniczna, lecz również fragmenty praktyk językowych, programistycznych, interpretacyjnych, klasyfikacyjnych i komunikacyjnych.

Fetyszyzm AI ukrywa systemową pracę ludzką

Ta perspektywa pozwala lepiej zrozumieć wykorzystaną przez Tenena marksowską metaforę "mądrego stołu". Autor nawiązuje tu do fragmentu Kapitału dotyczącego fetyszyzmu towarowego, przedstawiając technologiczny artefakt jako przedmiot, któremu przypisujemy niemal autonomiczne właściwości w momencie, gdy przestajemy dostrzegać społeczne relacje pracy stojące za jego powstaniem. Fetyszyzm w sensie marksowskim nie jest jednak prostym, przesądnym uwielbieniem przedmiotów; to forma społecznej percepcji, w której stosunki między ludźmi przyjmują wygląd stosunków między rzeczami.

W przypadku współczesnej AI analogiczny mechanizm zachodzi wtedy, gdy stwierdzenie "algorytm podjął decyzję" przesłania wcześniejsze wybory dotyczące zbierania danych, architektury systemu, funkcji celu, progu klasyfikacji, procedury wdrożenia czy organizacyjnego sposobu wykorzystania wyniku. Personifikacja maszyny może służyć jako skrót językowy, lecz w wymiarze politycznym bywa on znaczący – rozprasza odpowiedzialność między tak wiele elementów infrastruktury, że zaczyna ona sprawiać wrażenie, jakby nie należała do nikogo.

Koncepcja forensic labor attribution zasługuje na potraktowanie jako samodzielna hipoteza badawcza. Tenen postuluje "kryminalistyczną atrybucję pracy" w odpowiedzi na proces, w którym ludzkie doświadczenie zostaje przekształcone w zapis, następnie w korpus, a ostatecznie w materiał treningowy – przy czym wkład poszczególnych osób zanika wewnątrz zagregowanej struktury modelu. W notatce przykładem jest przejście od żywego doświadczenia pacjenta do transkrypcji i dokumentacji wykorzystywanej później w automatyzowanych systemach diagnostycznych. Pojęcie to można rozszerzyć poza problem autorstwa literackiego.

Dotyczy ono całego łańcucha pracy niezbędnego do zbudowania systemu: autorów źródeł, programistów, anotatorów, moderatorów treści, testerów bezpieczeństwa, osób tworzących odpowiedzi referencyjne, użytkowników dostarczających sygnały zwrotne oraz pracowników utrzymujących infrastrukturę danych. Analiza ekonomiczna AI powinna zatem obejmować nie tylko koszt GPU i energii, lecz również wielowarstwową strukturę pracy ukrytej za produktem przedstawianym jako niemal całkowicie automatyczny. Literatura empiryczna potwierdza, że intuicja Tenena nie jest wyłącznie metaforą humanistyczną.

Międzynarodowa Organizacja Pracy opisuje rozbudowany sektor pracowników zajmujących się oznaczaniem, kategoryzacją i weryfikacją danych, moderacją czy transkrypcją. Praca ta często organizowana jest przez platformy crowdworkowe lub przedsiębiorstwa outsourcingowe AI-BPO, co wiąże się z niskimi płacami, niestabilnością zatrudnienia oraz ograniczoną ochroną socjalną. ILO zwraca również uwagę na długotrwałą ekspozycję moderatorów danych treningowych na treści przedstawiające przemoc i nienawiść. Z polityczno-ekonomicznego punktu widzenia mamy zatem do czynienia z paradoksem: narracja o automatyzacji opiera się na systematycznym ukrywaniu pracy wykonywanej właśnie po to, aby produkt mógł wyglądać na autonomiczny.

Problem ten można analizować również przez pryzmat ekonomii kosztów transakcyjnych Ronalda Coase'a. Coase wykazał, że istnienie przedsiębiorstwa wynika m.in. z kosztów korzystania z rynku: wyszukiwania kontrahentów, negocjowania i monitorowania umów. Firma powstaje tam, gdzie koordynacja administracyjna jest tańsza niż seria transakcji rynkowych. Generatywna AI może przesuwać te granice.

Jeśli koszt przygotowania analizy, tłumaczenia, kodu czy dokumentacji gwałtownie maleje, przedsiębiorstwo może ograniczyć outsourcing i wykonywać te zadania wewnętrznie. Z drugiej strony modele udostępniane jako usługa mogą wzmacniać outsourcing – firma nie buduje wtedy własnej kompetencji, lecz kupuje ją w chmurze od wyspecjalizowanego dostawcy.

Nie ma więc jednego kierunku wpływu AI na granice przedsiębiorstwa; technologia zmienia jedynie względne koszty alternatywnych sposobów organizacji pracy. Rozwinięta przez Olivera Williamsona ekonomia kosztów transakcyjnych pozwala dodać do tej analizy problem specyficzności aktywów, niepewności i mechanizmów governance. Dane korporacyjne i modele dostrojone do konkretnych procesów mogą mieć wysoką wartość wewnętrzną, lecz niewielką poza danym środowiskiem. Jednocześnie uzależnienie od jednego dostawcy API lub platformy chmurowej tworzy nową postać zależności kontraktowej i technologicznego lock-in. W rezultacie polityczna ekonomia AI nie sprowadza się do pytania o liczbę zaoszczędzonych godzin pracy.

Obejmuje ona pytanie, czy kompetencja przenosi się z pracownika do przedsiębiorstwa, z przedsiębiorstwa do dostawcy infrastruktury, czy też powstaje układ wzajemnej zależności, w którym podmiot posiadający model przejmuje strategiczną pozycję wobec organizacji niezdolnych samodzielnie odtworzyć tej kompetencji. Jeszcze inną perspektywę daje współczesna ekonomia zadań Darona Acemoglu i Pascuala Restrepo.

Automatyzacja zadań a nie zawodów zmienia strukturę pracy

W tym ujęciu technologia nie „automatyzuje zawodu” jako niepodzielnej całości, lecz zmienia podział zadań między pracę a kapitał. Automatyzacja generuje efekt wypierania w momentach, gdy zadanie dotychczas wykonywane przez człowieka zostaje przejęte przez system techniczny lub kapitał. Jednocześnie wzrost produktywności może stymulować popyt na pracę w obszarach odpornych na automatyzację, a powstawanie nowych zadań może ponownie zwiększyć udział człowieka w procesie produkcji. Taka teoria znacznie lepiej oddaje empiryczne realia generatywnej AI niż popularna dychotomia: „zawód pozostanie albo zniknie”. Prawnik jako profesja może nadal istnieć, mimo że research, pierwszy szkic pisma czy analiza dokumentów zostaną częściowo zautomatyzowane; podobnie lekarz zachowa centralną rolę, choć systemy cyfrowe przejmą transkrypcję wizyt, kodowanie dokumentacji i przygotowanie części diagnozy różnicowej. Aktualne dane ILO potwierdzają zasadność tego zadaniowego podejścia.

W indeksie opublikowanym w maju 2025 roku około jedna czwarta pracowników na świecie zajmowała stanowiska wykazujące pewien stopień ekspozycji na generatywną AI, jednak tylko 3,3 procent światowego zatrudnienia przypadało na najwyższą kategorię ekspozycji. Zjawisko to było silniejsze w gospodarkach wysokodochodowych i szczególnie widoczne w zawodach biurowych, przy czym rozwój modeli zwiększył ekspozycję również w części profesji technicznych i wysoko kwalifikowanych. Kluczowy wniosek ILO nie dotyczy jednak masowego znikania miejsc pracy, lecz transformacji struktury zadań – większość zawodów bowiem obejmuje czynności wymagające dalszego udziału człowieka. W tym miejscu teza Tenena o automatyzacji knowledge work wymaga doprecyzowania: automatyzacja rzeczywiście dotarła do lekarzy, prawników, pisarzy i programistów, ale nie oznacza to zautomatyzowania całych zawodów. Materiał źródłowy wskazuje na szerokie spektrum zadań – od moderacji i wyszukiwania, po dokumentację medyczną i analizę informacji – a nie na jednolitą eliminację profesjonalisty.

Znaczenie automatyzacji dla pracy zależy ponadto od kierunku projektowania technologii. Acemoglu i Restrepo rozróżniają ścieżki koncentrujące się na zastępowaniu pracowników od rozwiązań tworzących nowe zadania, w których praca ludzka posiada przewagę komparatywną. Ich krytyka „niewłaściwego rodzaju AI" dotyczy sytuacji, w której innowacja skupia się nadmiernie na automatyzacji istniejących czynności, zamiast rozszerzać produktywność człowieka i otwierać nowe możliwości działania. Jest to ekonomiczny odpowiednik sporu Kirchera i Kuhlmanna, przywołanego wcześniej w tym artykule.

Narzędzie może zastępować etap procesu albo działać jak poznawcze rusztowanie zwiększające kompetencję użytkownika. Efekt społeczny zależy zatem od projektu organizacyjnego, systemu wynagradzania, procedur szkolenia i podziału korzyści produktywnościowych, a nie od samego faktu istnienia algorytmu.

Polityczna ekonomia AI łączy się tutaj z teorią algorytmicznego zarządzania. ILO definiuje algorithmic management jako wykorzystywanie systemów opartych na śledzonych danych do organizowania, przydzielania, monitorowania, nadzorowania i oceniania pracy. Nie każdy taki system musi korzystać z zaawansowanej AI; kluczowe jest samo przeniesienie części funkcji menedżerskich do systemu obliczeniowego. Konsekwencje tego procesu są ambiwalentne.

Automatyzacja zarządzania może zwiększać spójność decyzji, usprawniać harmonogramy i obniżać koszty koordynacji. Może jednak również intensyfikować nadzór nad pracownikiem, ograniczać jego autonomię, potęgować presję czasu i czynić podstawę decyzji trudną do zakwestionowania. Badanie OECD obejmujące ponad sześć tysięcy firm w sześciu krajach wskazuje właśnie na współistnienie oczekiwanych zysków efektywnościowych z problemami dotyczącymi zaufania i odpowiedzialnego governance.

Prawo i ekonomia w obliczu automatyzacji wiedzy

Europejskie prawo pracy zaczyna odpowiadać na tę transformację w sposób, który dobrze ilustruje Tenenowską tezę, że technologia "koduje politykę". Dyrektywa UE 2024/2831 dotycząca pracy platformowej wprost ustanawia cele związane z przejrzystością, sprawiedliwością, nadzorem ludzkim, bezpieczeństwem i odpowiedzialnością w algorithmic management; część jej reguł dotyczących przetwarzania danych obejmuje również osoby wykonujące pracę platformową bez formalnego stosunku pracy. Państwa członkowskie mają czas na jej transpozycję do 2 grudnia 2026 roku.

Jest to prawnie doniosłe przesunięcie, gdyż system algorytmiczny przestaje być traktowany jedynie jako prywatne narzędzie techniczne przedsiębiorstwa. Sposób, w jaki przydziela on pracę, monitoruje wykonawcę i wpływa na warunki świadczenia usług, staje się przedmiotem regulacji praw podmiotowych. Drugim zasadniczym problemem politycznej ekonomii modeli generatywnych jest własność intelektualna. Debata publiczna często upraszcza ten problem do pytania, czy "AI kradnie książki", lecz konstrukcja prawna jest bardziej złożona. W Unii Europejskiej art. 4 dyrektywy DSM 2019/790 ustanawia wyjątek lub ograniczenie dotyczące text and data mining dla legalnie dostępnych utworów, pod warunkiem że uprawniony nie zastrzegł odpowiednio korzystania z dzieła; w przypadku treści udostępnianych online zastrzeżenie może przyjmować postać odczytywalna maszynowo. Akt w sprawie sztucznej inteligencji nie zastępuje tego systemu prawa autorskiego, lecz nakłada na dostawców modeli ogólnego przeznaczenia obowiązek wdrożenia polityki zgodności z unijnym prawem autorskim – w tym respektowania zastrzeżeń praw – oraz obowiązek publikowania wystarczająco szczegółowego streszczenia treści wykorzystanych do treningu.

Tenenowskie forensic labor attribution zderza się tutaj z prawną konstrukcją autorstwa i praw wyłącznych. Prawo autorskie identyfikuje konkretne utwory, twórców i uprawnienia; model statystyczny powstaje natomiast z gigantycznej liczby źródeł, a wpływ pojedynczego tekstu na konkretną późniejszą odpowiedź bywa trudny do śledzenia.

Obowiązujący od 2 sierpnia 2025 roku reżim dla general-purpose AI w UE próbuje częściowo odpowiedzieć na tę asymetrię poprzez obowiązek streszczenia danych treningowych. Komisja wskazuje, że publiczny opis ma ułatwiać podmiotom posiadającym uzasadniony interes, w tym właścicielom praw autorskich, wykonywanie ich praw. Nie jest to jednak prawny odpowiednik pełnej księgi wkładów każdego autora i nie tworzy automatycznie prawa do wynagrodzenia za każdy fakt wykorzystania dzieła w procesie treningowym. Tenenowska postulowana atrybucja pracy jest więc szerszym programem normatywnym niż aktualne obowiązki prawne. Znaczenie tej

rozbieżności rośnie, gdy uwzględnimy ekonomię dóbr niematerialnych: tekst, kod czy wzorzec stylistyczny nie zużywają się fizycznie w taki sposób jak materiał w produkcji przemysłowej, a jednocześnie prawa wyłączne, licencje i kontrola dostępu determinują możliwość ich ekonomicznego wykorzystania.

Model generatywny komplikuje tradycyjną relację między kopią, transformacją a produktem pochodnym, ponieważ nie działa po prostu jako katalog fragmentów źródłowych, choć w szczególnych warunkach może reprodukować treści zapamiętane. Z punktu widzenia ekonomii instytucjonalnej zasadniczym pytaniem jest zatem sposób zdefiniowania praw, które równoważą koszt tworzenia dóbr kultury, społeczną wartość dostępu do wiedzy i możliwość prowadzenia badań z ryzykiem, że renty z automatyzacji zostaną nadmiernie skoncentrowane w podmiotach kontrolujących infrastrukturę obliczeniową. Tenenowska teza o pracy zbiorowej komplikuje również prosty język własności. Jeżeli model jest rezultatem zbiorowej historii tekstów i praktyk, nie wynika z tego automatycznie, że każdy uczestnik tej historii posiada prawo własności do modelu. Prawo własności i prawo autorskie są instytucjami normatywnymi, a nie opisem metafizycznego pochodzenia rzeczy. Można jednak postawić inne, ekonomicznie ważniejsze pytanie: czy podział korzyści odpowiada podziałowi wkładów, ryzyka i kosztów społecznych? To pytanie prowadzi ku teorii rent ekonomicznych i koncentracji rynku.

Trenowanie największych modeli wymaga obecnie olbrzymiej infrastruktury obliczeniowej, dostępu do kapitału, centrów danych, chipów, energii oraz zbiorów danych. Bariery wejścia mogą sprawiać, że technologia umożliwiająca demokratyzację pojedynczej kompetencji równocześnie koncentruje kontrolę nad infrastrukturą tej kompetencji. Ekonomia przedsiębiorstwa zna podobne paradoksy z wcześniejszych technologii sieciowych: wartość dla użytkownika może rosnąć wraz ze skalą usługi, natomiast wysokie koszty stałe i niskie koszty krańcowe dostarczania kolejnej jednostki mogą sprzyjać koncentracji. AI dodaje do tego efekt uczenia się z użytkowania, dostęp do danych oraz integrację pionową pomiędzy chmurą, modelami a aplikacjami. W rezultacie konkurencja nie musi przebiegać tylko między "lepszymi modelami", lecz może dotyczyć całych stosów infrastrukturalnych obejmujących procesory, centra danych, modele bazowe, dystrybucję i aplikacje końcowe. Tenenowska "mówiąca biblioteka" jest więc nie tylko biblioteką kultury.

Odpowiedzialność prawna AI to odpowiedzialność konkretnych podmiotów ludzkich

System AI jest własnością konkretnej organizacji lub usługą świadczoną na podstawie określonego kontraktu, przez co kwestia odpowiedzialności prawnej jest nierozerwalnie związana z ekonomią tej

technologii. W języku potocznym system może "odmówić kredytu", "zdiagnozować chorobę" czy "zwolnić kierowcę z platformy", jednak prawo musi zrekonstruować łańcuch podmiotów i ich obowiązków. Tenen słusznie ostrzega w swoich dziewięciu tezach końcowych, że personifikowanie algorytmów może zdejmować odpowiedzialność z twórców, właścicieli oraz organizacji wdrażających technologię. Stan europejskiego prawa w sierpniu 2026 roku wskazuje na próbę odejścia od metafizyki "autonomicznego podmiotu AI" na rzecz przypisania obowiązków konkretnym uczestnikom procesu. Od 2 sierpnia 2026 roku Komisja oraz organy krajowe rozpoczęły egzekwowanie kolejnych zasad AI Act, w tym nowych wymogów przejrzystości – między innymi w zakresie informowania użytkownika o kontakcie z systemem AI oraz oznaczania treści syntetycznych.

Konstrukcja regulacyjna nie przyznaje modelowi osobowości prawnej; zamiast tego dystrybuuje obowiązki między dostawców, wdrażających i inne podmioty uczestniczące w cyklu życia systemu. Jeszcze wyraźniej widać to w prawie odpowiedzialności za produkt. Nowa dyrektywa UE 2024/2853 wprost włącza oprogramowanie, w tym systemy AI, do pojęcia produktu dla celów odpowiedzialności za produkt wadliwy, uwzględniając możliwość powstawania defektów w związku z aktualizacjami oraz algorytmami uczenia maszynowego pozostającymi pod kontrolą producenta. Należy jednak precyzyjnie zaznaczyć aspekt temporalny: dyrektywa ma zastosowanie do produktów wprowadzonych do obrotu lub oddanych do użytku po 9 grudnia 2026 roku. Oznacza to, że w chwili sporządzania tego artykułu – 29 sierpnia 2026 roku – nowy reżim nie obejmuje jeszcze tak określonych przyszłych produktów.

Jednocześnie projekt odrębnej AI Liability Directive został wycofany w 2025 roku. Europejski system odpowiedzialności za szkody wyrządzone przez AI pozostaje zatem wielowarstwową konstrukcją, złożoną z AI Act, prawa produktowego, przepisów krajowych oraz reżimów ochrony danych i regulacji sektorowych, a nie jednolitym kodeksem odpowiedzialności algorytmicznej. Z punktu widzenia filozofii prawa jest to istotne: odpowiedzialność nie wymaga przypisania systemowi świadomości ani wolnej woli. System może być przyczynowo istotnym elementem szkody, nie będąc podmiotem moralnym. Prawo od wieków potrafi oddzielać przyczynę materialną od podmiotu odpowiedzialnego normatywnie – wadliwe urządzenie może spowodować szkodę, lecz zobowiązanie odszkodowawcze przypisuje się producentowi, operatorowi lub innemu uczestnikowi stosunku prawnego. Problem AI jest trudniejszy ze względu na adaptacyjność, złożoność i rozproszenie łańcucha dostaw, ale nie wymusza on automatycznie kreowania "osoby elektronicznej". Tenenowskie porównanie AI do korporacji lub Lewiatana okazuje się w tej perspektywie bardziej płodne niż figura samotnego robota. Korporacja również funkcjonuje w systemie społecznym dzięki sieci ludzi, dokumentów i procedur oraz prawnych fikcji organizacyjnych, mimo braku jednego biologicznego mózgu obejmującego całość jej wiedzy.

Hobbesowski trop wymaga rozwinięcia. Lewiatan był sztuczną osobą polityczną zbudowaną przez agregację działań jednostek, a jego siła wynikała właśnie z instytucjonalnego skupienia rozproszonej sprawczości.

AI jako narzędzie reorganizacji władzy i odpowiedzialności

Podobnie nowoczesne przedsiębiorstwo dysponuje wiedzą wykraczającą poza kompetencje któregokolwiek z pracowników, posiada pamięć rozciągniętą na archiwa i bazy danych oraz zdolność działania, która trwa mimo wymiany personelu. Kiedy Tenen pisze, że AI "pamięta jak rodzina, myśli jak państwo, rozumie jak korporacja", nie należy traktować tego jako tezy z zakresu neuronauki. To analogia socjologiczna: pewne formy inteligencji mogą być właściwością zorganizowanego systemu działania, choć nie dają się przypisać pojedynczemu członkowi tej struktury. W tym kontekście materiał źródłowy przeciwstawia indywidualistycznej metafizyce inteligencji pojęcie kolektywnego i rozproszonego systemu poznawczego.

Taka interpretacja nie zdejmuje jednak z nas pytania o władzę. Rozproszone poznanie opisuje sposób, w jaki kompetencja rozmieszcza się pomiędzy ludźmi a artefaktami; polityczna ekonomia pyta natomiast, kto kontroluje warunki tej dystrybucji. Pracownik może korzystać z modelu jako rozszerzenia własnego poznania, lecz ten sam model może służyć organizacji do monitorowania jego pracy, pomiaru produktywności i ograniczania swobody decyzyjnej. Technologia potrafi demokratyzować dostęp do analizy prawnej i jednocześnie wzmacniać pozycję podmiotów kontrolujących infrastrukturę. Może podnosić wydajność jednoosobowego przedsiębiorcy, a zarazem zwiększać koncentrację rynku usług chmurowych. Pomaga lekarzowi w diagnozie, lecz równocześnie przekształca wiedzę medyczną w zasób kontrolowany przez zamkniętą platformę.

Interdyscyplinarna analiza AI musi zatem unikać zarówno determinizmu technologicznego, według którego technologia sama dyktuje rezultaty społeczne, jak i naiwnego instrumentalizmu, zakładającego, że jest ona neutralnym narzędziem, niezależnym od sposobu swojego zaprojektowania. W tym świetle Tenenowska zasada "technology encodes politics" zyskuje najpełniejsze rozwinięcie. Architektura techniczna nie determinuje jednej polityki, lecz wyznacza koszty, możliwości i asymetrie, w ramach których operują później polityka oraz gospodarka. System wymagający ogromnych nakładów obliczeniowych tworzy bowiem inne warunki koncentracji niż mały model uruchamiany lokalnie.

Model pozbawiony mechanizmu cytowania generuje inne możliwości kontroli źródeł niż system oparty na jawnej atrybucji. Platforma, której ranking automatycznie determinuje dostęp pracownika do zleceń, tworzy inny układ sił niż narzędzie pozostawiające decyzję koordynatorowi podlegającemu negocjowanym regułom. Prawo, rynek i projekt techniczny są zatem współzależnymi instytucjami. Z tej perspektywy dziewięć wielkich idei Tenena należy czytać jako program badań nad polityczną ekonomią wiedzy, a nie antytechnologiczny manifest. "Collective labor" prowadzi do pytania o dystrybucję korzyści; "Distributed intelligence" do kwestii granic organizacji i odpowiedzialności. Automatyzacja knowledge work kieruje nas ku teorii zadań, kapitału ludzkiego i rekonfiguracji profesji, natomiast "Technology encodes politics" do analizy governance i praw pracowniczych.

Z kolei "Metaphors obscure responsibility" odnosi się do prawa odpowiedzialności, a postulat forensic labor attribution prowadzi do problemu autorstwa, praw wyłącznych i widzialności wkładów społecznych. Tenenowska humanistyka maszyn okazuje się dzięki temu nie dodatkiem do ekonomii AI, lecz sposobem na ujawnienie pytań, które czysto techniczny opis systemu pozostawia poza kadrem. Kluczowe jest tu porzucenie dychotomii "człowiek kontra maszyna". Realne konflikty XXI wieku znacznie częściej przebiegają między różnymi sposobami organizowania współpracy ludzi z technologią: między pracownikiem używającym AI a przedsiębiorstwem wykorzystującym ją do nadzoru, między autorem a właścicielem infrastruktury treningowej, między użytkownikiem a dostawcą modelu, czy między społecznym interesem w dostępie do wiedzy a ekonomiczną koniecznością finansowania jej wytwarzania. To spór między produktywnością a autonomią oraz między techniczną możliwością działania a normatywnym prawem do działania.

Model nie jest więc po prostu nowym "pracownikiem" wchodzącym na rynek, lecz technologią reorganizującą stosunki między istniejącymi podmiotami. Polityczna ekonomia sztucznego intelektu musi zatem kończyć się pytaniem o podmiotowość, choć nie w jej potocznym znaczeniu. Nie chodzi o to, czy model posiada świadomość, lecz o to, komu system społeczny przypisuje prawo do podejmowania decyzji i ich wykonywania, zdolność czerpania korzyści oraz obowiązek ponoszenia konsekwencji. W tej perspektywie filozofia moralna, prawo i teoria polityczna wracają do centrum wywodu.

Jeżeli Tenen ma rację, że metafora "inteligentnej maszyny" może przesłaniać sieć ludzi i instytucji, kolejnym zadaniem jest ustalenie, czy sztuczna inteligencja może być podmiotem moralnym, czy też jej pozorna podmiotowość jest przede wszystkim szczególną formą delegowania sprawstwa przez człowieka, korporację i państwo.

AI jako techne bez phronesis i charakteru moralnego

Właśnie ten problem wyznacza kolejny etap naszej analizy: przejście od woli do odpowiedzialności oraz przyjrzenie się filozoficznym tradycjom podmiotowości moralnej w obliczu sztucznej inteligencji.

Teza Dennisa Yi Tenena, według której współczesne maszyny nie powinny być traktowane jako podmioty moralne, nabiera właściwego znaczenia dopiero wtedy, gdy osadzimy ją w wielowiekowej historii pojęć sprawstwa, woli, odpowiedzialności, rozumu praktycznego i osoby. Argument Tenena wiąże się przede wszystkim z biologiczną i egzystencjalną kondycją człowieka: algorytmy nie doświadczają bólu, strachu przed śmiercią ani miłości. Zatem nie uczestniczą w tej wspólnocie wzajemnej podatności na krzywdę, z której wyrastają ludzkie praktyki moralne.

Jednocześnie autor porównuje współczesne systemy do Hobbesowskich „Lewiatanów", sugerując, że ich pozorna autonomiczność może przesłaniać realne działanie korporacji i instytucji. Z kolei spór Kuhlmanna z Kircherem służy mu jako historyczna figura różnicy między poprawnym wykonaniem procedury a „wewnętrznym rozumieniem". Jest to stanowisko istotne, lecz nie wolno przedstawiać go jako niekwestionowanego rezultatu współczesnej etyki. Historia myśli moralnej pokazuje bowiem, że różne tradycje lokowały podstawę odpowiedzialności w różnych przymiotach: w dobrowolności działania, zdolności rozumnego wyboru, zgodzie woli, autonomii, charakterze, uczuciach moralnych, reakcji na racje, relacji uznania czy uczestnictwie w społecznych praktykach przypisywania winy i zasługi. Dopiero zderzenie tych tradycji pozwala ocenić, które z tych warunków mogą, a których nie mogą spełniać obecne systemy AI.

U źródeł zachodniej teorii odpowiedzialności stoi Arystoteles. Jego etyka nie jest teorią pojedynczych poprawnych decyzji, lecz koncepcją dobrego życia – eudaimonia – realizowanego poprzez trwałe dyspozycje charakteru (aretai), rozwijane w praktyce. Człowiek moralnie dojrzały nie jest urządzeniem maksymalizującym poprawność wyników. To istota, której sposób postrzegania sytuacji, emocje, pragnienia i rozum praktyczny ukształtowały się tak, by potrafiła rozpoznać, co w konkretnych okolicznościach jest właściwe. Arystotelesowska phronesis – roztropność lub mądrość praktyczna – jest czymś więcej niż znajomością reguł. Pozwala ona zastosować normy w rzeczywistości, w której przypadki nie są identyczne, a etyczna złożoność wymyka się ogólnej instrukcji. Arystoteles analizuje również warunki przypisywania pochwały i nagany, odróżniając działania dobrowolne od tych wynikających z przymusu lub niewiedzy. Możliwość wskazania podmiotu jako źródła czynu stała się w ten sposób kluczowa dla późniejszych koncepcji odpowiedzialności.

Z perspektywy współczesnej AI arystotelesowskie rozróżnienie między techne a phronesis nabiera szczególnego znaczenia. Techne to umiejętność wytwarzania zgodnie z celem i regułami sztuki, podczas gdy phronesis dotyczy praktycznego rozumowania o tym, jak żyć i działać dla ludzkiego dobra w zmiennych sytuacjach. System może doskonale wykonywać operacje techniczne, a mimo to nie posiadać charakteru moralnego w sensie arystotelesowskim, gdyż charakter jest rezultatem habituacji całej osoby; organizuje on nie tylko wnioskowanie, ale i pożądania, reakcje emocjonalne oraz sposób przeżywania dobra. Tu pojawia się głębsza wersja krytyki Tenena: nawet perfekcyjny klasyfikator moralnych przypadków nie staje się automatycznie istotą cnotliwą, tak jak podręcznik medycyny nie staje się lekarzem tylko dlatego, że zawiera poprawne zalecenia. Cnota jest dyspozycją podmiotu do właściwego działania, a nie jedynie właściwością wygenerowanego rezultatu.

Filozoficzne kryteria sprawstwa a symulacja intencjonalności

Stoicy przesunęli środek ciężkości ku temu, co dzieje się między pojawieniem się wrażenia a jego przyjęciem przez rozumny podmiot. Kluczowym pojęciem stała się sunkatathesis – zgoda udzielana wrażeniu. Człowiek nie kontroluje wszystkich zdarzeń ani nawet pierwszych reakcji psychicznych, lecz może dokonać krytycznej oceny tego, co mu się jawi, i przyjąć lub zawiesić zgodę. Stanford Encyclopedia of Philosophy podkreśla, że właśnie akty zgody były przez stoików traktowane jako szczególny przedmiot odpowiedzialności moralnej i epistemicznej: nie odpowiadamy za samo pojawienie się impresji, lecz za sposób, w jaki rozumna część duszy się wobec niej ustosunkowuje. Ten stoicki motyw okaże się zdumiewająco trwały.

W różnych formach powraca on w chrześcijańskiej teorii woli, nowożytnym pojęciu autonomii oraz współczesnej filozofii odpowiedzialności, która często pyta nie o to, czy podmiot kontrolował wszystkie przyczyny swojej sytuacji, lecz czy posiadał zdolność refleksyjnego odnoszenia się do motywów własnego działania. Święty Augustyn nadał problemowi głęboko chrześcijańską postać, przesuwając odpowiedzialność jeszcze wyraźniej ku woli i wewnętrznej zgodzie. W jego wczesnej teorii bodziec zewnętrzny może wywołać wrażenie i skłonność, lecz dopiero wewnętrzne przyzwolenie tworzy wolę prowadzącą do moralnie znaczącego czynu. Z tego powodu człowiek może być odpowiedzialny za zamiar nawet wtedy, gdy zewnętrzne działanie nie powiedzie się; odwrotnie zaś fizyczne zdarzenie dokonujące się bez zgody podmiotu nie jest jego moralnym czynem. Późny Augustyn skomplikuje tę teorię nauką o upadłej naturze i łasce: wola pozostaje centrum odpowiedzialności, lecz jej zdolność do wyboru dobra okazuje się ograniczona przez

ludzką kondycję. Dla genealogii odpowiedzialności jest to moment doniosły, ponieważ podmiot moralny zostaje określony nie przez sprawność wykonawczą, ale przez zdolność wewnętrznego przyzwolenia, zamiaru i odnoszenia własnego działania do dobra.

Augustyńska tradycja stanowi wymagające kryterium dla sztucznej inteligencji. System generatywny może wytworzyć zdanie „chcę pomóc", lecz fenomen językowy nie rozstrzyga, czy istnieje podmiot, dla którego dobro drugiego człowieka jest rzeczywistym przedmiotem woli. Symulacja form gramatycznych intencjonalności nie jest dowodem istnienia Augustyńskiej voluntas. Problem ten nie może zostać rozwiązany przez analizę samego tekstu wyjściowego, ponieważ argument dotyczy rodzaju wewnętrznego związku między podmiotem, motywem a działaniem. Tutaj Tenenowskie „inward understanding" Kuhlmanna wpisuje się w znacznie starszą linię filozoficzną: rezultat zachowania nie wyczerpuje pojęcia moralnej podmiotowości.

Tomasz z Akwinu przeprowadził najbardziej systematyczną średniowieczną syntezę Arystotelesowskiej teleologii i chrześcijańskiej teorii woli. W jego ujęciu wola jest racjonalnym pożądaniem, a człowiek działa jako podmiot moralny wtedy, gdy może deliberować nad różnymi środkami prowadzącymi do rozpoznanego dobra. Akwinata odróżniał zwykłe zdarzenia zachodzące w człowieku od actus humanus – czynu ludzkiego we właściwym znaczeniu, którego zasadniczą cechą jest dobrowolność i rozumne samookreślenie. Jego etyka opiera się na idei dominium sui actus, panowania człowieka nad własnym czynem; bez takiego samookreślenia pojęcia zasługi, winy i powinności traciłyby podstawę. Akwinata dostarcza przy tym rozróżnienia niezwykle użytecznego dla filozofii AI.

Czym innym jest działanie rzeczy zmierzającej ku celowi zgodnie z porządkiem nadanym jej przez zewnętrzną przyczynę, a czym innym działanie podmiotu, który sam poznaje cel i kieruje ku niemu własne działanie. Współczesny system optymalizacyjny posiada funkcję celu w technicznym sensie, lecz fakt ten nie rozstrzyga, czy posiada cel w tomistycznym sensie intencjonalnym. Funkcja straty może organizować zachowanie sieci, ale nie wynika z tego, że sieć rozpoznaje dobro i wybiera je jako dobro. To rozróżnienie staje się szczególnie ważne w epoce agentic AI, gdzie słowo „cel" stosuje się jednocześnie do matematycznego kryterium optymalizacji, celu nadanego przez użytkownika oraz rzekomej własnej intencji systemu. Filozofia średniowieczna uczy, że tych trzech znaczeń nie należy mieszać.

Nowożytność wprowadziła do problemu sprawstwa kolejne napięcie. Thomas Hobbes przyjął mechanistyczną koncepcję natury i nie potrzebował libertariańskiej wolności woli, aby mówić o odpowiedzialności. Dla naszej genealogii jeszcze ważniejsza jest jednak jego teoria osoby sztucznej. W Lewiatanie osoba może reprezentować działania innych dzięki procesowi autoryzacji; państwo jest sztuczną osobą stworzoną przez ludzi, którzy ustanawiają strukturę reprezentacji i władzy.

Hobbesowska osoba polityczna nie jest biologicznym organizmem, lecz posiada zdolność działania w porządku społecznym dzięki instytucjonalnemu przypisaniu autorstwa czynów. Właśnie dlatego Tenenowska metafora AI jako „Lewiatana" staje się bardziej interesująca, jeśli nie interpretuje się jej jako twierdzenia, że komputer posiada duszę państwa, lecz jako wskazanie na możliwość organizowania sprawstwa przez systemy przekraczające jednostkę.

Hobbes pozwala zarazem dostrzec zasadniczą różnicę między osobą naturalną a sztuczną oraz między sprawczością a odpowiedzialnością. Korporacja może podpisywać umowy, posiadać majątek i ponosić określone rodzaje odpowiedzialności dlatego, że porządek instytucjonalny ustanawia reguły przypisywania jej działań; nie musi przy tym doświadczać bólu jako organizm. Tradycja osoby prawnej komplikuje zatem biologiczny argument Tenena. Jeżeli pytanie brzmi, czy AI może kiedykolwiek stać się podmiotem prawnym, brak biologicznego cierpienia sam w sobie nie daje rozstrzygającej odpowiedzi, ponieważ prawo od dawna posługuje się podmiotami niebiologicznymi.

Brak autonomii moralnej i zdolności do wzajemnego uznania

Jeśli jednak pytanie dotyczy odpowiedzialności moralnej w sensie winy, skruchy, intencji i zasługi, argumentacja wymaga innych kryteriów. Należy tutaj zatem oddzielić podmiotowość prawną od moralnej. David Hume przesunął teorię moralności jeszcze dalej od racjonalistycznego obrazu człowieka jako czystego kalkulatora norm. Jego słynne twierdzenie, że rozum sam w sobie nie jest źródłem motywacji moralnej, wiąże się z teorią uczuć moralnych. Aprobata i dezaprobata wyrastają z ludzkiej zdolności do przeżywania reakcji emocjonalnych wobec charakteru i działań innych, co jest następnie korygowane przez szerszą perspektywę. Hume łączył moralność z sympatią, uczuciami i społecznym doświadczeniem obserwatora, a nie z dedukcją norm z czystego rozumu. Z tej perspektywy Tenenowskie podkreślanie bólu, miłości i biologicznej podatności zyskuje silniejsze historyczne zaplecze: jeśli moralność jest nierozerwalnie związana z określonym rodzajem afektywnej reakcji na dobro i cierpienie innych, to system zdolny jedynie do klasyfikowania opisanych emocji nie staje się przez to uczestnikiem Hume'owskiej wspólnoty uczuć. Jednocześnie Hume wskazuje na interesującą trudność.

Uczucia moralne nie są wyłącznie prywatnymi reakcjami fizjologicznymi, lecz podlegają społecznej korekcie poprzez przyjmowanie wspólnego punktu widzenia. Moralność posiada zatem wymiar intersubiektywny. To sprawia, że problem AI nie można sprowadzić do pytania o to, czy maszyna „czuje" w sensie fenomenalnym. Trzeba również pytać, czy może ona uczestniczyć w praktyce

wzajemnego korygowania ocen, rozumieć społeczną pozycję drugiego podmiotu i odnosić własne działania do jego interesów w sposób, który nie byłby jedynie instrumentalnym wykonywaniem zewnętrznej funkcji.

Immanuel Kant stworzył wpływową racjonalistyczną alternatywę dla sentymentalizmu. W centrum jego filozofii moralnej leży autonomia: istota moralna nie tylko podlega prawu, ale jako rozumny podmiot nadaje sobie prawo moralne. Imperatyw kategoryczny ma obowiązywać bezwarunkowo, a osoba jest czymś więcej niż środkiem do realizacji cudzych celów właśnie dlatego, że posiada zdolność autonomicznego rozumu praktycznego. Kantowska etyka osłabia zatem tezę, jakoby moralność musiała bezpośrednio wyrastać z bólu, śmiertelności czy biologicznej potrzeby przetrwania. Dla Kanta źródłem zobowiązania jest racjonalna autonomia, a nie cierpienie. Gdyby kiedyś powstał system rzeczywiście zdolny do autonomicznego stanowienia i uznawania norm za wiążące racje działania, sama jego niebiologiczność nie wystarczyłaby do wykluczenia go z Kantowskiego królestwa celów. Warunek ten jest jednak znacznie silniejszy niż zdolność generowania języka moralnego. Model może napisać rozprawę o imperatywie kategorycznym, sklasyfikować dylemat albo przewidzieć społeczną ocenę czynu, nie będąc przez tę normę w żaden sposób zobowiązanym. Kantowskie „uznanie racji" nie jest tym samym co statystyczna reprezentacja zdania zawierającego rację.

Autonomia wymaga, aby źródłem działania była zdolność podmiotu do samodzielnego związania się prawem, a nie wyłącznie reakcja systemu na zewnętrzny prompt, funkcję nagrody czy politykę dostawcy. To rozróżnienie stanowi jeden z najmocniejszych argumentów przeciwko pochopnemu utożsamianiu dzisiejszego agenta programowego z agentem moralnym. Hegel komplikuje ten obraz poprzez teorię uznania.

Autonomia nie jest u niego wyłącznie własnością izolowanego, rozumnego „ja". Podmiotowość rozwija się społecznie: człowiek rozpoznaje siebie jako wolny podmiot w relacji z innymi, których również musi uznać za wolnych. W dojrzałej Filozofii prawa indywidualna wolność urzeczywistnia się dopiero w ramach instytucjonalnych porządków rodziny, społeczeństwa obywatelskiego i państwa. Współczesne badania nad Heglowską koncepcją recognition podkreślają właśnie tę wzajemność: adekwatne uznanie wymaga relacji między podmiotami zdolnymi traktować się nawzajem jako autonomiczne źródła racji. Z perspektywy Hegla pytanie o AI zmienia charakter. Nie chodzi jedynie o to, czy wewnątrz urządzenia znajduje się odpowiedni mechanizm poznawczy, lecz czy system może wejść w relację wzajemnego uznania, w której rozumie samego siebie jako podmiot normatywny i uznaje drugiego nie tylko za zmienną w funkcji celu, lecz za autonomicznego partnera. Obecne modele LLM potrafią znakomicie symulować język uznania,

przeprosin, zobowiązania czy szacunku, jednak symulacja praktyk językowych pozostaje czymś innym niż uczestnictwo w Heglowskiej formie życia etycznego.

Iluzja intencjonalności a rzeczywista odpowiedzialność moralna

Jednocześnie teoria Hegla ostrzega przed błędem skrajnego indywidualizmu: odpowiedzialność osób korzystających z AI jest osadzona w instytucjach, a nie wyłącznie w psychice pojedynczego użytkownika. Moralna ocena systemów musi zatem uwzględniać strukturę organizacji, rynku i prawa.

Dwudziestowieczna filozofia działania nadała problemowi intencji znacznie większą precyzję. Elizabeth Anscombe w pracy Intention wykazała, że działanie intencjonalne rozpoznajemy m.in. przez szczególną rolę pytania „dlaczego?". Człowiek działający intencjonalnie potrafi ująć swój czyn pod określonym opisem i wskazać rację, dla której go wykonuje. Donald Davidson rozwinął tę myśl w przyczynowej teorii działania, według której czyn jest intencjonalny pod danym opisem, gdy zostaje odpowiednio wyjaśniony przez „primary reason" – zazwyczaj kombinację przekonania i pragnienia. Ta tradycja dobitnie pokazuje, dlaczego słowo „agent", używane w informatyce, jest filozoficznie wieloznaczne.

Program może wykonywać działania celowe w sensie funkcjonalnym, lecz filozofia działania pyta, czy jego zachowanie posiada strukturę intencjonalności właściwą działaniu „z racji", a nie jedynie przyczynowe ukierunkowanie na rezultat. Współczesne systemy potrafią oczywiście wygenerować odpowiedź na pytanie „dlaczego to zrobiłeś?", jednak nie oznacza to, że ta racja była faktyczną przyczyną działania. W interpretacji AI niezwykle łatwo ulec błędowi retrospektywnej racjonalizacji: system najpierw generuje wynik, a dopiero później produkuje przekonujące wyjaśnienie zgodne z konwencjami dyskursu.

Dla teorii odpowiedzialności różnica ta jest zasadnicza. Podmiot odpowiada za działanie dlatego, że określone racje rzeczywiście kierowały jego decyzją; samo dostarczenie narracyjnego uzasadnienia post factum nie wystarcza. Właśnie dlatego explainable AI nie może być pojmowane jedynie jako zdolność modelu do opowiadania o własnym procesie. Wyjaśnienie powinno pozostawać epistemicznie związane z rzeczywistym mechanizmem, który doprowadził do decyzji.

Harry Frankfurt przesunął dyskusję o odpowiedzialności z pytania „czy można było zrobić inaczej?" w stronę identyfikacji podmiotu z motywem własnego działania. Jego słynne kontrprzykłady miały podważyć Principle of Alternative Possibilities, zgodnie z którym odpowiedzialność zawsze wymaga

realnej możliwości postąpienia inaczej. Frankfurtowska linia teorii woli zwraca uwagę na hierarchiczne relacje między pragnieniami: dojrzały podmiot nie tylko czegoś chce, ale może aprobować lub odrzucać własne pragnienia, kształtując wolę, z którą się identyfikuje.

Dla sztucznej inteligencji jest to kryterium niezwykle wymagające. Optymalizacja celu pierwszego rzędu nie wystarcza; należałoby wykazać zdolność systemu do refleksyjnego odnoszenia się do własnych motywów i uznawania jednych z nich za takie, które powinny go konstytuować.

Inny zwrot przyniósł Peter Strawson. W pracy Freedom and Resentment z 1962 roku powiązał on odpowiedzialność z siecią interpersonalnych „reactive attitudes": wdzięczności, urazy, przebaczenia, oburzenia czy poczucia winy. Zamiast szukać metafizycznego rozwiązania problemu determinizmu, Strawson wskazał na praktykę życia społecznego, w której traktujemy ludzi jako podmioty zobowiązań i oczekiwań. Możemy przejściowo przyjąć wobec kogoś „objective attitude", traktując jego zachowanie raczej jak zjawisko wymagające zarządzania lub terapii niż czyn adresowany do nas przez równorzędnego partnera moralnego. Jednak to właśnie perspektywa wzajemnych oczekiwań stanowi zasadniczy wymiar ludzkich relacji.

Strawsonowski test jest szczególnie interesujący w kontekście chatbotów, gdyż ludzie spontanicznie zaczynają stosować wobec nich język reaktywnych postaw: dziękują im, obrażają się, przypisują im arogancję, uczciwość, życzliwość lub manipulację. Jest to istotny fakt psychologiczny, lecz nie dowód moralnej podmiotowości systemu. Strawson analizował praktykę międzyosobową ukształtowaną przez oczekiwania dobrej woli innych ludzi. Jeśli jedna strona relacji jedynie generuje wzorce odpowiadające językowi przeprosin i troski, nie wiadomo, czy powstała symetryczna relacja moralna, czy jedynie antropomorfizacja niezwykle skutecznego interfejsu. Tenenowska metafora „mówiącej biblioteki" może tutaj służyć jako korekta intuicyjnego błędu: responsywność językowa wystarcza do wywołania naszych reaktywnych postaw, lecz nie przesądza o istnieniu odpowiedniego podmiotu po drugiej stronie interakcji.

Na koniec Bernard Williams i Thomas Nagel wprowadzili do teorii odpowiedzialności problem moral luck. Nasze intuicje sugerują, że człowiek powinien odpowiadać tylko za to, co pozostaje pod jego kontrolą, jednak realne oceny moralne często zależą od czynników przypadkowych. Dwaj równie nieostrożni kierowcy mogą zostać ocenieni zupełnie inaczej, jeśli przed samochodem tylko jednego z nich znajdzie się pieszy. Nagel wyróżniał m.in. szczęście dotyczące rezultatu, okoliczności, konstytucji podmiotu oraz causal luck.

Luka odpowiedzialności nie uzasadnia personifikacji AI

Problem ten nabiera szczególnego znaczenia w systemach AI, gdzie wielość zależności technologicznych dodatkowo komplikuje relację między kontrolą a skutkiem. Programista może nie przewidzieć konkretnego zachowania modelu, operatorowi może zabraknąć czasu na skuteczną interwencję, a użytkownik może nie znać ograniczeń systemu – mimo to powstała szkoda wymaga społecznego przypisania odpowiedzialności. Stąd wyrasta współczesny problem responsibility gap. Andreas Matthias argumentował już w 2004 roku, że adaptacyjne i uczące się systemy mogą generować zachowania, których producenci i operatorzy nie byli w stanie szczegółowo przewidzieć, co sprawia, że klasyczne wzorce przypisywania odpowiedzialności stają się trudne do zastosowania. Nie należy jednak wnioskować, że luka ta powinna zostać automatycznie wypełniona moralną odpowiedzialnością maszyny. Alternatywne strategie obejmują odpowiedzialność za projekt instytucjonalny, wdrożenie systemu mimo niepewności, obowiązek monitorowania, zasady odpowiedzialności produktowej, ubezpieczenie ryzyka czy rozproszone przypisanie obowiązków wielu uczestnikom łańcucha. Problem luki Matthiasa jest zatem pytaniem o adekwatność dotychczasowych mechanizmów governance, a nie dowodem na narodziny syntetycznej osoby moralnej. Madeleine Clare Elish rozwinęła ten problem z przeciwnej strony, wprowadzając koncepcję moral crumple zone.

W wysoko zautomatyzowanym układzie człowiek znajdujący się najbliżej punktu awarii może absorbować winę za porażkę całego systemu, mimo że jego faktyczna kontrola była niewielka. Elish analizuje ten mechanizm jako odpowiednik strefy kontrolowanego zgniotu w samochodzie: odpowiedzialność moralna i prawna koncentruje się na operatorze, podczas gdy strukturalne decyzje projektantów, przedsiębiorstw i instytucji pozostają niewidoczne. Koncepcja ta niezwykle dobrze uzupełnia Tenena.

Personifikowanie maszyny może rozmywać odpowiedzialność organizacji, lecz równie niebezpieczne bywa deklarowanie, że "człowiek pozostaje w pętli", jeśli w rzeczywistości nie posiada on już czasu, informacji ani realnej władzy niezbędnej do skorygowania decyzji automatycznego systemu. Dyskusja o odpowiedzialności grupowej prowadzi jeszcze dalej. Christian List i Philip Pettit argumentują, że odpowiednio zorganizowane grupy mogą spełniać funkcjonalne warunki agencyjności, posiadając mechanizmy tworzenia przekonań, celów i decyzji, których nie można utożsamić z przekonaniami pojedynczego członka. Teorie collective intentionality i group agency pokazują, że odpowiedzialność nie musi zawsze kończyć się na jednostce biologicznej, choć trwa spór między stanowiskami indywidualistycznymi a kolektywistycznymi o to, czy kolektyw może być rzeczywistym podmiotem odpowiedzialności. To ponownie wzmacnia instytucjonalne odczytanie Tenena.

System AI może funkcjonować jako element agenta korporacyjnego, nie będąc samodzielnym agentem moralnym. Firma posiada cele, procedury decyzyjne, role, kanały korekty i prawnie uznane mechanizmy działania; model staje się wówczas częścią tej architektury, podobnie jak baza danych, zarząd, dział prawny czy procedura audytu. Luciano Floridi i J. W. Sanders zaproponowali w tym kontekście bardziej radykalne podejście.

Rozróżnienie między funkcjonalną sprawczością a świadomym podmiotem

W artykule On the Morality of Artificial Agents argumentowano, że na odpowiednio dobranym poziomie abstrakcji można mówić o sztucznych agentach moralnych, nie wymagając od nich wszystkich klasycznych przymiotów ludzkiej osoby, takich jak wolna wola czy pełna odpowiedzialność w tradycyjnym sensie. Autorzy rozdzielają tam moralną agencyjność od odpowiedzialności, proponując analizę systemów przez pryzmat interaktywności, autonomii i adaptacyjności. Stanowi to istotny kontrargument wobec stanowiska Tenena: pojęcie „moral agent" może mieć bowiem techniczny, funkcjonalny sens, szerszy niż określenie podmiotu zdolnego do winy, skruchy czy zasługi.

Ceną za takie podejście jest jednak wyraźne odejście od rozumienia moralnego sprawcy, jakie rozwijali Arystoteles, Augustyn, Kant czy Strawson. Spór okazuje się więc częściowo sporem terminologicznym o to, jakie minimum chcemy zawrzeć w pojęciu agencyjności moralnej.

Warto wprowadzić tutaj rozróżnienie często pomijane w dyskusjach popularnych: między moral agency a moral patiency. Podmiot moralny w pierwszym sensie jest tym, komu można przypisać obowiązki i odpowiedzialność; moral patient zaś to byt, którego dobro samo w sobie może nakładać na innych moralne ograniczenia. Niemowlę, osoba w głębokim zaburzeniu świadomości czy wiele zwierząt posiada status moralnego pacjenta mimo braku pełnej odpowiedzialności moralnej. Dlatego nawet jeśli przyszłe systemy AI nie staną się autonomicznymi sprawcami, pozostaje pytanie, czy posiadają one doświadczenia, interesy lub zdolność cierpienia uzasadniającą obowiązki wobec nich.

W sierpniu 2026 roku nauka nie daje podstaw do kategorycznego stwierdzenia, że współczesne modele językowe są świadome. Problem pozostaje otwarty metodologicznie: zespół badaczy świadomości, neuronauki i filozofii pod kierunkiem Patricka Butlina proponuje ocenianie systemów za pomocą wskaźników wyprowadzanych z konkurencyjnych teorii naukowych, podkreślając przy tym znaczną niepewność zarówno samych teorii, jak i metod pomiaru. Nie

istnieje więc naukowa podstawa, by samą płynność dialogu czy deklarację „odczuwam ból” uznać za dowód doświadczenia fenomenalnego. Trzeba odróżnić behawioralną kompetencję językową od ontologicznego pytania, czy istnieje "coś, jak to jest" być danym systemem.

Zestawienie różnych tradycji prowadzi do obrazu bardziej złożonego niż prosty spór o to, czy maszyna posiada moralność. Z Arystotelesa otrzymujemy wymóg charakteru, habituacji i praktycznej mądrości; od stoików – zdolność refleksyjnego przyzwolenia; od Augustyna – wagę woli i wewnętrznej zgody; a od Akwinaty – samookreślenie racjonalnej woli i świadome odnoszenie działania do celu. Hobbes wskazuje na możliwość sztucznej podmiotowości instytucjonalnej, Hume wprowadza uczucia moralne i społeczną sympatię, Kant autonomię i samozobowiązanie wobec prawa moralnego, zaś Hegel wzajemne uznanie i społeczno-instytucjonalny charakter wolności.

Współczesna myśl kontynuuje te rozważania: Anscombe i Davidson analizują intencję oraz działanie z racji, Frankfurt pyta o refleksyjną identyfikację z własnymi motywami, a Strawson osadza odpowiedzialność w interpersonalnej praktyce oczekiwań i reaktywnych postaw. Z kolei Williams i Nagel wskazują na granice kontroli, podczas gdy List i Pettit rozszerzają dyskusję na podmiotowość grupową.

Rozróżnienie sprawstwa AI chroni przed rozmyciem odpowiedzialności instytucjonalnej

Matthias wskazuje na lukę odpowiedzialności wynikającą z nieprzewidywalności systemów uczących się, Elish na ryzyko przerzucania winy na najbliższego człowieka, natomiast Floridi i Sanders kwestionują konieczność zachowania wszystkich antropologicznych warunków przy definiowaniu sztucznego agenta moralnego. Na tle tych rozważań stanowisko Tenena można sformułować znacznie precyzyjniej niż w potocznej tezie, że „maszyna nie czuje, więc nie jest moralna”. Jego argument należy rozumieć jako obronę bogatej, relacyjnej i ucieleśnionej koncepcji sprawstwa, według której uczestnictwo w moralności wymaga czegoś więcej niż tylko zdolność do osiągania społecznie aprobowanych wyników. Materiał źródłowy wiąże ten brak z nieobecnością biologicznych imperatywów: cierpienia, troski o przetrwanie i „wewnętrznego rozumienia”. Historia filozofii pokazuje jednak, że biologiczność jest tylko jedną z możliwych podstaw tej argumentacji. Kant postawiłby problem autonomii, Arystoteles charakteru i phronesis, Anscombe wiedzy praktycznej, Strawson interpersonalnej odpowiedzialności, Hegel uznania, a współczesna teoria działania – zdolności reagowania na racje.

Obecne modele nie wykazują przekonująco pełnej postaci żadnego z tych zestawów własności. W konsekwencji najbezpieczniejszym i filozoficznie uzasadnionym podejściem do współczesnych systemów nie jest przypisywanie im pełnej odpowiedzialności moralnej, lecz wielopoziomowa analiza sprawstwa. System AI niewątpliwie posiada sprawczość przyczynową, gdyż jego działanie zmienia świat. Może posiadać sprawczość funkcjonalną, jeśli elastycznie realizuje cele i reaguje na zmiany środowiska. Może również uczestniczyć w sprawstwie instytucjonalnym jako element korporacji, administracji czy zespołu.

Nie oznacza to jednak posiadania autonomicznego sprawstwa moralnego, które pozwalałoby obciążyć system winą w sensie, w jakim obciążamy rozumnego człowieka. Tym bardziej nie wynika z tego automatycznie świadomość fenomenalna ani status moralnego pacjenta. Rozdzielenie tych kategorii jest jednym z najważniejszych wkładów, jakie filozofia może wnieść do debaty technologicznej. Pozwala ono zrozumieć, dlaczego język personifikujący AI ma realne konsekwencje polityczne. Gdy mówimy, że „algorytm odmówił", „AI zwolniła", „model zdecydował" lub „system popełnił błąd", niezauważalnie przechodzimy od opisu sprawstwa przyczynowego do sugestii sprawstwa normatywnego.

Tenen obawia się właśnie tego przesunięcia: w materiale źródłowym maszyna jest instrumentem w rękach konkretnych podmiotów ekonomicznych, a metafora autonomicznego aktora może maskować to, kto podjął decyzję o jej zaprojektowaniu, zakupie, wdrożeniu i zakresie delegowanej władzy. Elish wskazuje na symetryczne niebezpieczeństwo: organizacja może przekazać systemowi rzeczywistą kontrolę, pozostawiając człowieka jedynie jako nominalnego gwaranta bezpieczeństwa, by następnie przerzucić na niego winę w chwili awarii. Odpowiedzialność może zatem zostać utracona zarówno przez nadmierne personifikowanie maszyny, jak i przez fikcyjne utrzymywanie człowieka „w pętli".

Dziedzictwem cywilizacyjnej tradycji namysłu nad wolą nie jest jedna definicja odpowiedzialności, lecz rozpoznanie, że wymaga ona odpowiednio zbudowanego związku między podmiotem, wiedzą, możliwością kontroli, intencją, racjami działania a skutkiem. Sztuczna inteligencja nie unieważnia tej tradycji; przeciwnie – zmusza do jej ponownego użycia z większą precyzją.

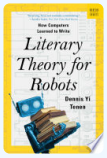
Po dwóch i pół tysiąca lat filozofia nie dostarcza nam prostego testu „czy maszyna jest osobą", ale oferuje aparat pozwalający rozłożyć to pytanie na elementy, które można badać osobno. Jest to znacznie cenniejsze niż efektowna odpowiedź metafizyczna, ponieważ na tych rozróżnieniach można budować realne zasady projektowania instytucji, prawa i odpowiedzialności. Dalszy krok prowadzi nieuchronnie od indywidualnej podmiotowości ku ustrojowi społecznemu, w którym używane są inteligentne narzędzia. Skoro odpowiedzialność wymaga rzeczywistej możliwości kontroli, instytucja nie może delegować operacyjnej władzy systemowi, a następnie pozostawiać

pracownika jako symboliczną "strefę zgniotu". Ponieważ autonomia jest warunkiem moralnego działania, technologia powinna być oceniana przez pryzmat tego, czy zwiększa, czy zmniejsza autonomię człowieka. Heglowskie uznanie wymaga traktowania drugiej osoby jako podmiotu, zatem automatyzacja administracyjna nie może redukować obywatela do profilu ryzyka bez mechanizmów zakwestionowania decyzji. Arystotelesowska phronesis wymaga osądu sytuacyjnego, więc system prawny i medyczny nie powinien utożsamiać wyników predykcyjnych z kompletnym rozstrzygnięciem normatywnym. Hume przypomina o moralnej roli współodczuwania – organizacja nie może zakładać, że poprawność statystyczna zastąpi relację troski. Wreszcie Kant wiąże godność z traktowaniem osoby jako celu; człowiek nie powinien stawać się jedynie źródłem danych, zmienną funkcji celu ani korektorem błędów cudzej infrastruktury.

Filozoficzna historia odpowiedzialności spotyka się tutaj z główną intuicją genealogii Tenena. Problemem sztucznej inteligencji może ostatecznie okazać się nie pytanie, czy udało się stworzyć drugi ludzki umysł, lecz to, jak techniczne rozszerzenia ludzkiego poznania zmieniają warunki, w których sami ludzie pozostają autorami i odpowiedzialnymi podmiotami własnych instytucji. Z tego problemu wyrasta konieczność analizy relacji między AI a prawem, demokracją, administracją, edukacją i konstrukcją Dobra Wspólnego – obszaru, w którym spór o inteligencję przestaje być domeną filozofii umysłu, a staje się sporem o model cywilizacji.

> Ostatecznie problem sztucznej inteligencji przestaje być zagadką z zakresu filozofii umysłu, a staje się fundamentalnym sporem o model naszej cywilizacji. Prawdziwym wyzwaniem nie jest zatem pytanie, czy udało nam się stworzyć syntetyczny ludzki intelekt, lecz to, w jakim stopniu techniczne rozszerzenia poznania pozwalają nam pozostać autorami własnego życia. Czy w świecie zdominowanym przez statystyczne prawdopodobieństwo i algorytmiczne zarządzanie znajdziemy jeszcze miejsce na autentyczną autonomię, która nie jest jedynie symulacją?

Inspiracje

[1] 

"Literary Theory for Robots" Dennis Yi Tenen

2024 | W. W. Norton | ISBN: 978-1324036173

Dennis Yi Tenen jest profesorem nadzwyczajnym języka angielskiego i literatury porównawczej na Uniwersytecie Columbia, gdzie współkieruje Centrum Mediów Porównawczych oraz Laboratorium Inteligencji Narracyjnej. Jego badania koncentrują się na styku ludzi, tekstu i technologii, obejmując takie dziedziny jak historia literatury, teoria mediów, humanistyka obliczeniowa i socjologia literatury. Przed rozpoczęciem kariery akademickiej Tenen pracował jako inżynier oprogramowania w firmie Microsoft w grupie Windows. Uzyskał doktorat z literatury porównawczej na Uniwersytecie Harvarda. Tenen, wieloletni współpracownik Instytutu Nauk o Danych Uniwersytetu Columbia i były pracownik naukowy Centrum Berkmana ds. Internetu i Społeczeństwa, znany jest z interdyscyplinarnej pracy nad historią inteligencji maszynowej i współpracą między autorami a inżynierami. Jest również założycielem Laboratorium Modelowania i Wizualizacji Literackiej Uniwersytetu Columbia i redaktorem serii wydawniczej „On Method".

Dennis Yi Tenen *is an associate professor of English and Comparative Literature at Columbia University, where he co-directs the Center for Comparative Media and the Narrative Intelligence Lab. His research explores the intersection of people, text, and technology, spanning fields such as literary history, media theory, computational humanities, and the sociology of literature. Before his academic career, Tenen worked as a software engineer at Microsoft in the Windows group. He holds a doctorate in Comparative Literature from Harvard University. A long-time affiliate of Columbia's Data Science Institute and a former fellow at the Berkman Center for Internet and Society, Tenen is known for his interdisciplinary work on the history of machine intelligence and the collaborative relationship between authors and engineers. He is also the founder of Columbia's Literary Modeling and Visualization Lab and edits the On Method book series.*

Columbia University

Literatura uzupełniająca

Książki

Autorzy przywoływani w artykule:

[1] "Teoria letteraria per robot. Come i computer hanno imparato a scrivere"
Dennis Yi Tenen
2024 | ISBN: 9788833943299

Literatura pokrewna (Google Scholar):

[2] "The Narrow Corridor"
Daron Acemoglu

[3] "Distorted innovation: Does the market get the direction of technology right?"
Daron Acemoglu

[4] "Leviathan"
Thomas Hobbes

[5] "Behemoth"
Thomas Hobbes

[6] "De Cive"
Thomas Hobbes

[7] "De iudiciis astrorum"
Tomasz Z Akwinu

[8] "Postilla super Psalmos"
Tomasz Z Akwinu

[9] "Super Boetium De Trinitate"
Tomasz Z Akwinu

[10] "Essays, Moral and Political"
David Hume

[11] "An Enquiry Concerning Human Understanding"
David Hume

[12] "A treatise of human nature"
David Hume

[13] **"View from nowhere"**

Thomas Nagel

[14] **"Mind and Cosmos"**

Thomas Nagel

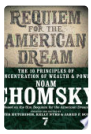
[15] **"An AI Pattern Language"**

Madeleine Clare Elish

[16] **"Morphology of the Folk Tale"**

Władimir Propp

Univ of TX + ORM | ISBN: 9780292748095



[17] **"Requiem for the American Dream"**

Noam Chomsky

Seven Stories Press | ISBN: 9781609807375

[18] **"Class Warfare"**

Noam Chomsky

Pluto Press (UK)

[19] **"The Fateful Triangle"**

Noam Chomsky

ISBN: 9781783712519

[20] **"semiology"**

Ferdinand De Saussure

[21] **"Cours de linguistique générale"**

Ferdinand De Saussure

BoD - Books on Demand | ISBN: 9782382749494

[22] **"Course in General Linguistics"**

Ferdinand De Saussure

Columbia University Press | ISBN: 9780231157278

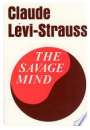
[23] **"The Naked Man"**

Claude Lévi-Strauss

University of Chicago Press | ISBN: 9780226474960

[24] **"La pensée sauvage"**

Claude Lévi-Strauss



[25] **"The Savage Mind"**

Claude Lévi-Strauss

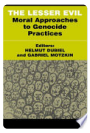
University of Chicago Press | ISBN: 9780226474847

[26] **"Programming a Computer for Playing Chess"**

Claude Shannon

[27] **"AI in Context: The Labor of Integrating New Technologies"**

Madeleine Clare Elish



[28] **"The Lesser Evil - Moral Approaches to Genocide Practices"**

Tzvetan Todorov

Routledge | ISBN: 9781135758813

[29] **"The Fantastic: A Structural Approach to a Literary Genre"**

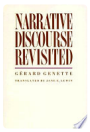
Tzvetan Todorov

Cornell University Press | ISBN: 9780801491467

[30] **"The Pleasure of the Text"**

Roland Barthes

Macmillan | ISBN: 9780374521608



[31] **"Narrative Discourse Revisited"**

Gérard Genette

Cornell University Press | ISBN: 9780801495359

[32] **"Modern Operating Systems"**

Tenenbaum

Prentice Hall | ISBN: 9780133591620

[33] **"Computer Networks"**

Tenenbaum

Pearson Education India | ISBN: 9788177581652

[34] **"Structured Computer Organization"**

Tenenbaum

Prentice Hall | ISBN: 9780132916523

[35] "Markov chain"

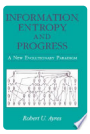
Andriej Markow

[36] "Markov process"

Andriej Markow

[37] "Markov property"

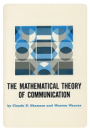
Andriej Markow



[38] "information entropy"

Claude Shannon

Springer Science & Business Media | ISBN: 9780883189115

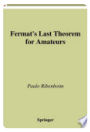


[39] "The mathematical theory of communication"

Claude Shannon

[40] "Theory of transaction costs"

Oliver Williamson



[41] "Fermat's Last Theorem"

Pierre De Fermat

Springer Science & Business Media | ISBN: 9780387216928

[42] "Fermat's little theorem"

Pierre De Fermat

[43] "Fermat number"

Pierre De Fermat

[44] "Bernoulli trial"

Jakob Bernoulli

[45] "Ars Conjectandi"

Jakob Bernoulli

[46] "binomial distribution"

Jakob Bernoulli

[47] **"Laplace's approximation"**

Pierre Simon De Laplace

[48] **"After the Ball"**

Lew Tołstoj

[49] **"Repentance"**

Lew Tołstoj

[50] **"War and Peace"**

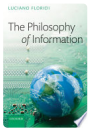
Lew Tołstoj

Signet Classics | ISBN: 9780451521163

[51] **"Information: A Very Short Introduction"**

Luciano Floridi

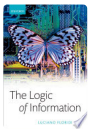
2010 | OUP Oxford | ISBN: 9780191609541



[52] **"The Philosophy of Information"**

Luciano Floridi

2013 | OUP Oxford | ISBN: 9780191655647



[53] **"The Logic of Information"**

Luciano Floridi

2019 | Oxford University Press | ISBN: 9780192570277

[54] **"vector space model"**

Gerard Salton

[55] **"Automatic Information Organization and Retrieval"**

Gerard Salton

New York : McGraw-Hill

[56] **"Skills, Tasks and Technologies: Implications for Employment and Earnings"**

Daron Acemoglu

[57] **"Dynamic Information and Library Processing"**

Gerard Salton

Prentice Hall

[58] **"The discovery of a world in the moone, or, A discourse tending to prove, that 'tis probable there may be another habitable world in that planet"**

John Wilkins

ISBN: 9782528103128

[59] **"Mathematical Magick"**

John Wilkins

[60] **"An Essay towards a Real Character and a Philosophical Language"**

John Wilkins

[61] **"Deep learning-enabled medical computer vision"**

Richard Socher

[62] **"Foundations of Statistical Natural Language Processing"**

Christopher Manning

[63] **"Stanford typed dependencies manual"**

Christopher Manning

[64] **"Introduction to Information Retrieval"**

Christopher Manning

[65] **"Human Problem Solving"**

Allen Newell

[66] **"Information Processing Language"**

Allen Newell

[67] **"Duncker on Thinking: an Inquiry Into Progress in Cognition"**

Allen Newell

[68] **"Organizations"**

Herbert Simon

[69] **"Models of my life"**

Herbert Simon

[70] **"Truth and Method"**

Hans-Georg Gadamer

[71] **"The Enigma of Health"**

Hans-Georg Gadamer

[72] **"The Idea of the Good in Platonic-Aristotelian Philosophy"**

Hans-Georg Gadamer

[73] **"Scaling Laws for Autoregressive Generative Modeling"**

Jared Kaplan

[74] **"On the Economy of Machinery and Manufactures"**

Charles Babbage

[75] **"Babbage's calculating engines"**

Charles Babbage

[76] **"Reflections on the Decline of Science in England, and on Some of Its Causes"**

Charles Babbage

[77] **"How China Became Capitalist"**

Ronald Coase

[78] **"The Problem of Social Cost"**

Ronald Harry Coase

1960

[79] **"Markets and Hierarchies : Analysis and Antitrust Implications,"**

Oliver Williamson

1975

[80] **"bounded rationality"**

Oliver Williamson

[81] **"Groundwork of the Metaphysics of Morals"**

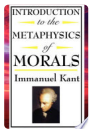
Immanuel Kant

2013 | CreateSpace | ISBN: 9781492204152

[82] **"Fundamental Principles of the Metaphysic of Morals"**

Immanuel Kant

2021 | Phoemixx Classics Ebooks | ISBN: 9783986477943



[83] "Introduction to the Metaphysics of Morals"

Immanuel Kant

2013 | Simon and Schuster | ISBN: 9781625584298

[84] "Human life, action and ethics: essays by G.E.M. Anscombe"

Elizabeth Anscombe

[85] "Presidential Addresses of the American Philosophical Association 1971-1980"

American Philosophical Association, Wilfrid Stalker Sellars, Arthur John Terence Dibben Wisdom, Henry Babcock Veatch, Lewis White Beck, Marjorie Glicksman Grene, May Selznick Brodbeck, Norman Adrian Malcolm, Patrick Suppes, Arthur Walter Burks, Donald Herbert Davidson, Alonzo Church, Alan Gewirth, John Bordley Rawls, Herbert Paul Grice, William Henry Hay, Alice Ambrose Lazerowitz, Kaarlo Jaakko Juhani Hintikka, Ruth Charlotte Barcan Marcus, Hilary Whitehall Putnam, Herbert Fingarette, Leonard Linsky, Kurt Erich Maria Baier, Wesley Charles Salmon, Robert George Turnbull, Monroe Curtis Beardsley, William Craig, William Payne Alston, Richard McKay Rorty, David Shepherd Nivison, Hector-Neri Castaneda

2013 | ISBN: 9780985974701

[86] "On Truth"

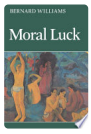
Harry Frankfurt

[87] "The Problem of Action"

Harry Frankfurt

[88] "On Bullshit"

Harry Frankfurt



[89] "Moral Luck"

Bernard Williams

1981 | Cambridge University Press | ISBN: 9781107268173

[90] "Moral Luck / Moralischer Zufall"

Bernard Williams

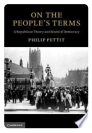
2025 | Reclam Verlag | ISBN: 9783159626079

[91] "The Last Word"

Thomas Nagel

[92] **"Assembling Accountability: Algorithmic Impact Assessment for the Public Interest"**

Madeleine Clare Elish



[93] **"On the People's Terms"**

Philip Pettit

2012 | Cambridge University Press | ISBN: 9781139851398

[94] **"The Common Mind"**

Philip Pettit

1993 | Oxford University Press, USA

[95] **"Human-Computer Interaction"**

Constantine Stephanidis, Gavriel Salvendy

2024 | CRC Press | ISBN: 9781040318607

[96] **"Mophology of the fairy tale. Disney's literary original "The Princess and the Frog" analysed on the basis of Propp's „Morphology of the folktale“"**

Kirsten Burmeister

2016 | GRIN Verlag | ISBN: 9783668254367

[97] **"Mophology of the Fairy Tale. Disney's Literary Original "The Princess and the Frog" Analysed on the Basis of Propp's "Morphology of the Folktale""**

Kirsten Burmeister

2016 | ISBN: 9783668254374

[98] **"From Fairy Tale to Fantasy"**

Maria Nikolajeva

1985

[99] **"A fairy tale's structure"**

Helga Mebus

2008 | GRIN Verlag | ISBN: 9783638065290

[100] **"Structural Analysis and the Concept of the "tale-type""**

Heda Jason

1968

Artykuły naukowe

Artykuły pokrewne (Google Scholar):

[1] "The Emergence of American Formalism"

Dennis Yi Tenen

2019 | Modern Philology | DOI: 10.1086/705672

Google Scholar

[2] "Distributed Agency in the Novel"

Dennis Yi Tenen

2022 | New Literary History | DOI: 10.1353/nlh.2022.a898333

Google Scholar

[3] "Toward a Computational Archaeology of Fictional Space"

Dennis Yi Tenen

2018 | New Literary History | DOI: 10.1353/nlh.2018.0005

Google Scholar

[4] "Author—Anonymous and Distributed"

Dennis Yi Tenen

2024 | New Techno-Humanities | DOI: 10.1016/j.techum.2024.07.001

Google Scholar

[5] "Materiality"

Dennis Yi Tenen

2024 | The Cambridge Companion to Literature in a Digital Age | DOI: 10.1017/9781009349567.009

Google Scholar

[6] "Reading Platforms"

Dennis Yi Tenen

2020 | The Unfinished Book | DOI: 10.1093/oxfordhb/9780198830801.013.22

Google Scholar

[7] "Archive"

Dennis Yi Tenen

2017 | Literature | DOI: 10.5040/9781474272001.ch-026

Google Scholar

[8] "What Is?: Nine Epistemological Essays"

Dennis Yi Tenen

2015 | Design and Culture | DOI: 10.1080/17547075.2015.1051841

Google Scholar

[9] "Blunt Instrumentalism:"

DENNIS TENEN

2016 | Debates in the Digital Humanities 2016 | DOI: 10.5749/j.ctt1cn6thb.12

Google Scholar

[10] "Stalin's PowerPoint"

Dennis Tenen

2014 | Modernism/modernity | DOI: 10.1353/mod.2014.0030

Google Scholar

Treść tworzy Fundacja Dobre Państwo.
Redaguje i publikuje APA ONE, autorski system redakcyjny Fundacji oparty na AI.
APA ONE Publishing System

Article ID: 3db61dc3-e3e0-45ce-a4a2-5a8f97f18c97